Lecture Notes in Mechanical Engineering

Jay Lee
Jun Ni
Jagnathan Sarangapani
Joseph Mathew *Editors*

Engineering Asset Management 2011

Proceedings of the Sixth World Congress on Engineering Asset Management

Springer

Lecture Notes in Mechanical Engineering

For further volumes:
http://www.springer.com/series/11236

Jay Lee · Jun Ni · Jagnathan Sarangapani
Joseph Mathew
Editors

Engineering Asset Management 2011

Proceedings of the Sixth World Congress on Engineering Asset Management

Springer

Editors
Jay Lee
Department of Mechanical, Industrial and Nuclear Engineering
College of Engineering
University of Cincinnati
Cincinnati, OH
USA

Jun Ni
Department of Mechanical Engineering
University of Michigan
Ann Arbor, MI
USA

Jagnathan Sarangapani
Missouri University of Technology
Rolla, MO
USA

Joseph Mathew
Centre for Integrated Engineering Asset Management
Queensland University of Technology
Brisbane, QLD
Australia

ISSN 2195-4356 ISSN 2195-4364 (electronic)
ISBN 978-1-4471-4992-7 ISBN 978-1-4471-4993-4 (eBook)
DOI 10.1007/978-1-4471-4993-4
Springer London Heidelberg New York Dordrecht

Library of Congress Control Number: 2013934026

Printed on acid-free paper

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

Contents

Chapter 1
Optimizing E-Maintenance Through Intelligent Data Processing Systems

E. Gilabert, E. Jantunen, C. Emmanouilidis, A. Starr and A. Arnaiz

Abstract The landscape of maintenance and asset management has been reshaped as key technology enablers that are making a significant impact on everyday applications. The growing maturing of web-based and semantic maintenance, the ubiquity of mobile and situated computing, and the lowered costs and increased capabilities of wireless sensing and identification technologies are among the enabling technologies having the most significant impact. They are recognized as the key constituents of eMaintenance, the technological framework that empowers organizations to streamline their asset management services and data delivery across the maintenance operations chain. This paper takes a look at these key, contributing technologies, alongside their adoption prospects and current hurdles preventing the wider penetration of eMaintenance in industry.

E. Gilabert (✉) · A. Arnaiz
Fundación Tekniker, Avda Otaola 20 20600 Eibar, Spain
e-mail: egilabert@tekniker.es

A. Arnaiz
e-mail: aarnaiz@tekniker.es

E. Jantunen
VTT Technical Research Centre of Finland, P.O.Box 1000 02044 VTT Espoo, Finland
e-mail: erkki.jantunen@vtt.fi

C. Emmanouilidis
ATHENA/CETI Research and Innovation Centre, 68100 Xanthi, Greece
e-mail: christosem@ceti.gr

A. Starr
Cranfield University, Cranfield MK43 0AL, UK
e-mail: a.starr@cranfield.ac.uk

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_1,

1.1 Introduction

During recent years, eMaintenance has become more and more popular. The rapid penetration of web-based, mobile, and wireless technologies in enterprise operations, alongside shrinking costs and increasing capacity of hardware, is changing the landscape of maintenance practice. At the same time, the manufacturers of machinery are strategically moving toward developing service business for taking care of their products throughout the entire life cycle of the equipment [1, 2]. In order to be able to do this, the manufacturers need tools for taking care of the machinery in a profitable and reliable way and this is where eMaintenance comes to the picture. eMaintenance can be considered a technology where information is provided where it is needed and maintenance is a task that is about information when done in an effective way. In modern maintenance scenarios, the actions are carried out at optimal timing before breakage and are based on need not on calendar. In recent years quite a lot of research has taken place covering various aspects of eMaintenance. The EU FP6 funded Dynamite project (Dynamic Decisions in Maintenance, IP017498) developed and tested a set of methodologies and tools to support the eMaintenance processes. The results of Dynamite are summarized in the recently published book on e-Maintenance [2]. The technological developments included smart tags, sensors, signal analysis, smart decision support, portable computing devices, maintenance web services, common database schemas, diagnosis, prognosis, as well as financial cost-efficiency assessment. It has been argued that such technological advancements are likely to provide a boost to modern industries in their pursuit of upgrading their overall efficiency in managing their assets [3, 4].

This paper tries to take this issue a little further and thus in the following paragraphs the future development of some of the key areas of eMaintenance are discussed and such aspects as identification technologies, wireless sensors, mobile devices, Internet, distributed computing, use of Internet and data, and diagnosis and prognosis technologies.

1.2 Wireless Sensing and Identification

Industry employs condition monitoring systems which are now rapidly adopting technology innovations. The most significant upgrades in the technical infrastructure of condition monitoring systems are related to their increased computational capacity and the increased versatility offered by the incorporation of different wireless protocols, as advanced mini and micro-scale components and RF are integrated within sensor boards. This empowers the sensing end of the condition monitoring system to offer increased computational and connectivity support, making it easier to integrate supporting logic and tools, such as novelty detection, diagnostics and prognostics, as well as enhancements in computerized maintenance management systems and remote services.

Wireless sensor networks offer easy and customizable deployment of several sophisticated agents of small form-factor, making them suitable for wireless condition monitoring with sensor-embedded intelligence [5, 6]. These monitoring units can act as agents, capable of hosting automated computational and data storing operations that scale from filtering and preprocessing, to anomaly detection and diagnostics. This has brought about a growing wave of in the manufacturing area, wherein more scalable installations of wireless condition monitoring systems begin to co-exist alongside their wired counterparts [7]. Based on wireless condition monitoring components and systems, maintenance service providers are now not limited to employing static monitoring infrastructures and solutions. Condition monitoring data and services can now become ubiquitously available to technical staff, via mobile and handheld devices, or remotely via the web, coupling to services residing directly in the sensing infrastructure. Increasingly, wireless condition monitoring is making headways to industrial practice, primarily in the area of (SHM) [8] and equipment or process monitoring [9].

The flexibility offered by wireless condition monitoring allows a multitude of customisable readings to be taking from measurement points, while simultaneously executing data-processing routines and algorithms at the sensor node level, or collectively, by a cluster of sensing nodes. The capacity to host such data-processing services at the sensor node or the wireless sensor network level has a direct impact on the nature of the monitored asset itself, embedding an increasing level of self-aware operation [10].

Alongside wireless sensing, asset identification is enabled by the usage of auto-identification technology, such as RFID and image tags. RFID usage upgrades the asset capability to store limited information locally, facilitating on-site information data storage and retrieval. Among the initial beneficiaries of RFID adoption was supply chain management [11]. With increased interoperability offered and strengthened by the adoption of electronic product control (EPC) standards, RFIDs have emerged as a natural link between the physical and the IT world, supporting the concept of the self-serving asset [12]. Integrating asset identification with wireless sensing offers a strong drive to contextualize maintenance data and services delivery, e.g., sensor readings are linked to a specific asset, which in turn operates under certain conditions and workload [13].

An emerging trend is related to merging sensing with asset identification, such as in Intel's WISP platform [14], a feature particularly relevant to e-maintenance. Furthermore, WISP supports energy efficiency, by adopting self-powering technologies, typically energy harvesting. A further push for the wider applicability of asset identification technology may be offered by tag-printing technology, similar to that of inkjet printing [15], while the potential to integrate in the same production process both sensors as well as RFID tags is promising. If the technology push is successful and commercially viable, enterprises will be able to produce their own tags to adapt to their ever-changing asset management needs, with only limited additional resources and no re-integration costs.

1.3 Mobile Devices and Distributed Computing

With the increasing usage of wireless technologies and smart phones, data related to CM and equipment maintenance can become ubiquitously available to personnel, through handheld devices, or on the internet by simply linking to relevant maintenance services, provided by remote servers or by the asset monitoring facilities. This new perspective has created a potential new market niche [16]. In fact, Cloud Computing is emerged as a commercial reality, related to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services [17].

Maintenance service providers can offer highly efficient and customizable software solutions. Increased interconnection allows seamless operation, exchanging data between middle and upper level software, such as CMMS or ERP, and various wireless sensing and input modules. This is the essence of what has been referred to as Mobile Maintenance Management [18]. In this sense, it is possible to use handheld devices, such as PDAs, as well as miniaturized sensor solutions within a more flexible and decentralized CM and maintenance management integrated environment [4, 19]. This environment involves the use of mobile devices to perform typical maintenance management tasks such as work order management, communication with the service centre, asset maintenance tactical planning, reporting work, availability and location, retrieving maintenance history or documentation etc. Although most of these services are offered by existing systems such as ERP or CMMS, the industrial user would much rather work with a simple device offering a limited but clear set of interaction interfaces to retrieve or enter maintenance data. Thus, the user is empowered to become a dynamic mobile actor, operating in a dynamic environment, carrying the capacity to perform maintenance tasks with the support of ubiquitously available maintenance-supporting IT tools [20, 21].

With maintenance personnel becoming involved as mobile actors in the asset management process, it has also become important to seek to tailor the offered services to the exact needs of each actor. These needs depend on the role and function of the personnel, but also on the specific circumstances of the maintenance service request. For example, upon receiving a specific maintenance task order, industrial staff would need to locate the asset on the shop floor, have access to its maintenance history, retrieve information about spare parts availability, or indeed perform a condition assessment audit with the help of the PDA and sensing modules. Tailoring the available services and information availability to the specific demands is a feature that has been sought at a premium and is often referred to as context-aware or situated computing. The notion of context has been linked with computing for many years, largely associated with computational linguistics. Since the mid 1990s, there has been increasing attention to the role that context might have in adaptive computing. Specifically, the interest focused on how computer applications can be adapted to match the requirements or needs of different situations and users. With the prevalence of service-oriented computing, adaptation capacity has become synonymous to adapting the offered services and content. Consequently a context-aware maintenance support system is expected to tailor each service to the apparent usage context. This is often perceived by users as a capacity to provide 'intelligent content' or 'intelligent services', often presented through 'intelligent interfaces'. Yet it is only following the deeper penetration of mobile and wireless technology into maintenance and engineering asset management practice that contextualized computing emerges as a significant element in modern and future maintenance management practice.

The use of augmented and virtual reality is also posing significant issues related to context, as the perceived experience in such applications critically depend on the successful user immersion in a contextually relevant situation. The mobile maintenance actor may employ different devices for different services, at different locations and at different times. The actors are using the mobile devices in different maintenance task contexts. A typical example is a mobile actor employing a PDA in relation to a CM task. The PDA can be employed in tandem with an identification scheme, such as a localization technique or radio frequency ID (RFID) tag, to identify the monitored asset and obtain a better understanding of the task in hand. The PDA can be employed to support diagnosis and prognosis. This implies higher demands in data availability and CPU power, not yet readily available in industrial PDAs. Part of the processing can be undertaken by smart sensor solutions, delegating some of the CM tasks to the lowest possible processing level, the machine level. The developed solution must strike the right balance between local processing and information exchange between devices, implying trade-offs between energy efficiency, processing capacity, and the quality of the maintenance decision support process. In the future, additional support may become available through more extensive use of virtual and augmented reality solutions, which can help to 'immerse' the mobile actor in almost-real life scenarios and 3D machinery models, providing more practical insight. However, these solutions are still in their infancy and not widely available in current industry.

1.4 Internet and Data

The internet is considered as a disruptive technology because of its impact in the world and business, enabling a distributed communication and data storage, in synergy with other communication technologies. For maintenance purposes, the internet has been crucial in the maintenance concept, along with web services, which are becoming a fundamental technology for performing eMaintenance. Web services allow the central processes in eMaintenance through the existence of distributed services to perform information analysis, management, order execution, even data capture. Computers attached to the Internet are able to interchange messages through web services, since the HTTP port used by web services is always open, even in the most restrictive firewall configurations. Furthermore, the semantic web is a worldwide project which intends to create an universal medium for information exchange by giving meaning (semantics) to the content of documents on the web, so that they are understandable by machines. The semantic web provides the required data in real time, supported by web services, in a future eMaintenance scenario [22, 23]. Nevertheless, the adoption of this new architecture is a slow process for different reasons. Usually, companies have a lot of information in paper format, and it is not always easy to make changes in the company culture. Also the adoption of a standard for data format is required, that would support products through their life cycle and the itegration of different technologies. The communication among distributed components needs a protocol that should be managed in a centralized way as this complexity makes the management of data a key issue.

The MIMOSA organization has been dedicated to developing and encouraging the adoption of open information standards for Operations and Maintenance in manufacturing, fleet, and facility environments for some decades. MIMOSA provides the standard Open Systems Architecture for Condition Based Maintenance (OSA-CBM), for information acquisition processes, developed to support interoperability through different CBM components. The standard Open Systems Architecture for Enterprise Application Integration (OSA-EAI), also provided by MIMOSA and complementary to OSA-CBM, was created to solve the problem of integrating different applications. OSA-EAI is essentially a large database composed of hundreds of tables. The MIMOSA definition covers different issues related to measurements, condition monitoring, diagnosis, prognosis, and management of maintenance work orders. However, the main drawback is the adoption of OSA-EAI in companies, which may be a difficult step if relational databases are not familiar. OSA-EAI is well documented and is quite easy to download and install, but it is not an easy step to start using a database in a logical way. Actually, the installation requires higher level effort in discussing how OSA-EAI should be used in order it to be an effective tool. But one must ask if this is a hopeless route, since it is necessary to understand 100,000 items in a standard simply for definition of concepts. Another problem is to keep the database up to date. The semantic data ultimately has a low level definition, which offers many degrees of freedom. In order to provide interoperability among components and software, more effort is needed to provide an useful and easy way to interchange data. Another very important set of standards is not related to specific task modeling, but instead to facilitate data integration among different tasks.

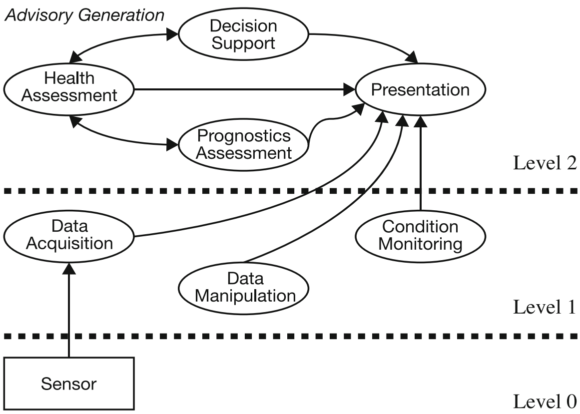
1.5 Diagnosis and Prognosis

The multisignal analysis and management is a challenge, that is, how to manage the data so that it is available for analysis and data from several sources can be integrated and also saved for later use for supporting the diagnosis and prognosis task. Usually, data is stored in hardware specific format even though MIMOSA aims for a common definition of data format, since several hardware solutions can support different analysis tools.

The transformation of maintenance strategy, and the escalation in the number of sensors and condition monitoring, will lead to pronounced need for automatic diagnosis, which also means that a massive leap forward has to be taken. The status of automatic diagnosis is still low in industry. For example, few consistent solutions are used in condition monitoring of rotating machinery, even though a lot of research has taken place in this field. It would seem that in many cases, the researchers have not had a wide enough view of the problem they have tried to solve. There are several studies about classification using, e.g., neural networks and consequently a great number of solutions have been developed which work in the laboratory with the type of data they have been trained with, but not in the field (see e.g. [23]) It would certainly seem that many researchers are naive if they assume that an ingenious classifier can solve the problem outright. The right kind of sensing of a physical phenomenon, and appropriate signal analysis, are essential—poor quality data will offer little chance of reliable classification. Many corporations are moving toward the provision of services for the machinery they produce, and much data will become accessible to support the condition monitoring task. Improved understanding of wear models for machines will especially help in the computerization of diagnosis. It is predictable that more process data will be accessible; simulations will be available and they make available supplementary information to support diagnosis and prognosis. When the improvement of signal analysis techniques is taken into account, the automatic diagnosis of the condition of machinery components can be anticipated to take a big step forward, and become more available for many kinds of components.

Additionally, the prognosis of failure of components suffering from wear as a function of loading will become possible on a level that supports the introduction of CBM. Due to the difficulty of diagnosis and prognosis of rotating machinery, data mining, and classification techniques as such do not really give a lot of support, but they will be very beneficial in handling spare parts, and reserving resources, based

Fig. 1.1 Correlation between OSA-CBM and ISO activity level (ISO 13374)



on more statistical than physical modeling. The ordering of spare parts and their administration will become semiautomatic. It is natural to start from the less costly parts that are used in numbers and then in time go to the more expensive and rare parts, where a human inspection process will probably be used for quite a long time in the future. Modern eMaintenance solutions could already carry out all the ordering and work force supervision automatically, but the potential for errors, e.g., sending out the wrong orders for parts and work is restricting this development. However, this process will naturally go further as more experience is gained from semiautomatic solutions (Fig. 1.1).

1.6 Conclusions

During the last decade, we have seen the progress of eMaintenance techniques. The key elements are the widespread use of Internet and quick progress of sensors and processing power. The use of eMaintenance at the moment still is at low level compared to the enormous potential but particularly inspired by the modifications of strategy of manufacturing industries toward the competence of providing services throughout the lifetime of the equipment they have manufactured will present a great increase in this technology. This step forward will be reinforced and supported by the hardware development but the most vital factor is the obtainability and usability of new data that can support diagnosis and prognosis and aid to raise them to a level that is of real assistance for the maintenance technicians. The advancement of new signal analysis techniques joint with simulation models can be the aspect that truly means a breakthrough in prognosis of the lifetime of components of machinery leading to the development of real proactive CBM. Many of the technologies are offered already now and their introduction could be straightforwardly vindicated economically, but obviously it will take some time to prioritize all the offered opportunities but naturally the accomplishment stories of those in the forefront will boost the adaption of new technologies in greater numbers.

References

1. Muller A, Crespo A, Iung B (2008) On the concept of e-maintenance: review and current research. J Relib Eng Syst Saf 93:1165–1187
2. Holmberg K, Adgar A, Arnaiz A, Jantunen E, Mascolo J, Mekid S (2010) E-maintenance. Springer, Berlin
3. Jantunen E, Gilabert E, Emmanouilidis C, Adgar A (2009) e-Maintenance: a means to high overall efficiency. In: Proceedings of 4th world congress on engineering asset management (WCEAM). Springer, Berlin

4. Liyanage JP, Lee J, Emmanouilidis C, Ni J (2009) Integrated e-Maintenance and intelligent maintenance systems. In: Ben-Daya M et al (eds) Handbook of maintenance management and engineering. Springer, Berlin, pp 499–544
5. Emmanouilidis C, Pistofidis P (2010) Wireless condition monitoring and embedded novelty detection. In: Amad-Echendu J et al (eds) Definitions, concepts and scope of engineering asset management, 1st edn. Springer, Berlin
6. Zhuang LQ, Goh, KM, Zhang JB (2007) The wireless sensor networks for factory automation: issues and challenges. In: IEEE conference on emerging technologies and factory automation (ETFA), IEEE. Patras, Greece, pp 141–148
7. Chintalapudi K, Fu T, Paek J, Kothari N, Rangwala S, Caffrey J, Govindan R, Johnson E, Masri S (2006) Monitoring civil structures with a wireless sensor network. IEEE Internet Comput 10:26–34
8. Alhetairshi M, Aramco S (2009) Vibration monitoring system using wireless technology: a feasible alternative solution for noncritical machines. In: Proceedings of SPE annual technical conference. New Orleans, Louisiana
9. Emmanouilidis C, Pistofidis P (2010) Machinery self-awareness with wireless sensor networks: a means to sustainable operation. In: Proceedings of the 2nd workshop on maintenance for sustainable manufacturing. Verona, Italy, pp 43–50
10. Angeles R (2005) RFID Technologies: supply-chain applications and implementation issues. J Inf Sys Manag 22:51–65
11. Brintrup A, Ranasinghe DC, Kwan S, Parlikad A, Owens K (2009) Roadmap to self-serving assets in civil aerospace. In: Proceedings of the 1st CIRP industrial product-service systems (IPS2) conference. Cranfield, UK, pp 323–330
12. Spaccapetra S, Al-Jadir L, Yu S (2005) Somebody, sometime, somewhere, something. In: Proceedings of 2005 workshop on ubiquitous data management. Tokyo, Japan, pp 6–16
13. Roy S, Jandhyala V, Smith JR, Wetherall DJ, Otis BP, Chakraborty R, Buettner M, Yeager D, Ko YC, Sample AP (2010) RFID: from supply chains to sensor nets. Proc of the IEEE. 98:1583–1592
14. http://www.seattle.intel-research.net/wisp
15. Lakafosis V, Vyas R, Nikolaou S, Tentzeris MM (2010) Progress towards the first wireless sensor networks consisting of inkjet-printed, paper based RFID enabled sensor tags. Proc of the IEEE 98:1601–1609
16. Jantunen E, Arnaiz A, Adgar A, Iung, B (2007) Mobile technologies for dynamic maintenance. In: Proceedings of the 3rd international conference on maintenance and facility management. Rome, Italy, pp 167–173
17. Armbrust M, Fox A, Griffith R, Katz RH, Konwinski A, Lee G, Patterson DA, Rabkin A, Stoika I, Zaharia M (2009) Above the clouds: a Berkeley view of cloud computing. UCB/EECS-2009-28
18. Arnaiz A, Emmanouilidis C, Iung B, Jantunen E (2006) Mobile maintenance management. J Int Technol Inf Manag 15:11–22
19. Emmanouilidis C, Liyanage JP, Jantunen E (2009) Mobile solutions for engineering asset and maintenance management. J Qual Maint Eng 15:92–105
20. Campos J, Jantunen E, Parkash O (2009) A web and mobile device architecture for mobile e-maintenance. Int J Adv Manuf Technol 45:71–80
21. Kopacsi S, Kovacs G, Anufriev A, Michelini R (2007) Ambient intelligence as enabling technology for modern business paradigms. J Robot Comput Integr Manuf 23:242–256
22. Gilabert E, Ferreiro S, Arnaiz A (2007) Web services system for distributed technology upgrade within an e-maintenance framework. On the move to meaningful internet systems 2007: OTM, workshops, pp 149–157
23. Fumagalli L, Jantunen E, Macchi M (2009) Ontology of mobile maintenance processes. In: Proceedings of MFPT 2009. Dayton, OH, USA

Chapter 2
Developing RCM Strategy for Wind Turbines Utilizing Online Condition E-Monitoring

D. Baglee and M. J. Knowles

Abstract The number of offshore wind turbines installed in the seas around Britain's coasts is likely to increase from just fewer than 150 to 7,500 over the next 10 years with the potential cost of £10 billion. Operation and Maintenance activities are estimated to comprise 35 % of the cost of electricity. However, the development of appropriate and efficient maintenance strategies is currently lacking in the wind industry. The current reliability and failure modes of offshore wind turbines are known and have been used to develop preventive and corrective maintenance strategies which have done little to improve reliability. Unplanned maintenance levels can be reduced by increasing the reliability of the gearbox and individual gears through the analysis of lubricants. In addition, the failure of one minor component can cause escalated damage to a major component, which can increase repair and or replacement costs. A Reliability Centered Maintenance (RCM) approach offers considerable benefit to the management of wind turbine operations, since it includes an appreciation of the impact of faults on operations. Due to the high costs involved in performing maintenance and the even higher costs associated with failures and subsequent downtime and repair, it is critical that the impacts are considered when maintenance is planned. The paper will provide an overview of the application of RCM and on line e-condition monitoring to wind turbine maintenance management. Finally, the paper will discuss the development of a complete sensor-based processing unit that can continuously monitor the wind turbine's lubricated systems and provide, via wireless technology, real-time data enabling onshore staff the ability to predict degradation, anticipate problems, and take remedial action before damage and failure occurs.

D. Baglee (✉) · M. J. Knowles
Institute for Automotive Manufacturing and Advanced Practices, University of Sunderland, Sunderland, UK
e-mail: david.baglee@sunderland.ac.uk

M. J. Knowles
e-mail: michael.knowles@sunderland.ac.uk

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_2,

2.1 Introduction

Wind power is becoming an increasingly important source of energy for countries that aim to reduce the emission of greenhouse gases and mitigate the effects of global warming. The wind power industry throughout Europe has experienced rapid growth especially within the United Kingdom (UK), where offshore development is predicted to increase steadily with the potential to generate 1,000 terawatt hours (TWh) per year, equivalent to several times the UK's total electricity consumption. In addition, worldwide wind power capacity has increased from 24,322 MW in 2001 to 159,213 MW in 2009 and the growth rate shows continued increase with capacity doubling every 3 years [1]. Failure rates have been found to be in the region of 1.5–4 failures per year while offshore turbines can need approximately five service/repair visits per year [2]. Having suitably trained personnel visit turbine installations is a costly enterprise, especially when large and heavy parts need to be replaced. Such problems are exacerbated in the case of offshore installations. Operational costs can be significantly reduced by developing an appropriate asset management strategy utilizing one or a number of modern maintenance technologies. The answer, for improved maintenance management and increased reliability, is to combine Reliability Centered Maintenance (RCM), e-condition monitoring and e-maintenance technologies. RCM is used due to its structured approach to find the appropriate maintenance tasks and requirements of assets in their specific operating context. This makes RCM particularly effective for identifying the maintenance requirements of equipment in the unique and demanding environment [3]. The use of RCM and e-condition monitoring for equipment maintenance is not widespread for wind turbines due to the difficulties in data transfer associated with their remote location. It is, however, a crucial aspect in terms of ensuring that environmentally and financially expensive maintenance visits are performed to maximum benefit and occur only when necessary. The costs and logistical implications of operatives visiting turbines means that decision making and planning of maintenance activities is critical for economic viability. Our proposition is that the provision of condition data gathered from e-condition sensors can be utilized optimally in conjunction with e-maintenance tools and a structured maintenance methodology to provide optimal cost savings. The rest of this paper is structured as follows. The following section provides a brief overview of the technologies involved in wind turbines. This is followed by an overview of the reliability issues related to the various components. New approaches to wind turbine maintenance utilizing RCM and e-maintenance solutions are discussed. The paper is concluded with conclusions regarding the efficient management of reliability of wind turbines.

2.2 Wind Turbine Technology

The drive to reduce costs has lead to rapid increases in turbine size. Small turbines are less economical in terms of land usage, maintenance requirements, and installed capacity. Figure 2.1 shows the increases that have occurred in turbine size while Fig. 2.2 shows the effect the average production cost associated with different sizes of turbines. Large-scale systems use asynchronous machines to generate electrical power and as such a gearbox is used to match the rotational speed of the machine to the frequency required by the power grid. Smaller domestic and commercial installations convert the generated alternating current to direct current before using an inverter to produce AC at the required frequency. The mechanical components are broadly common involving bearings and gear systems.

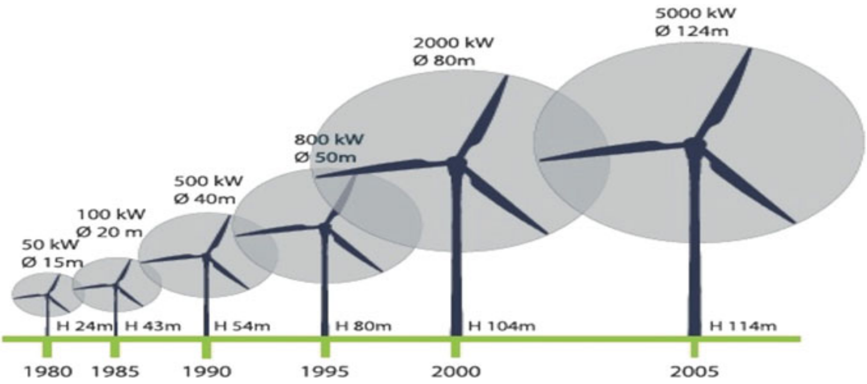


Fig. 2.1 Trend in increasing wind turbine size (*Source* EWEA 2007)

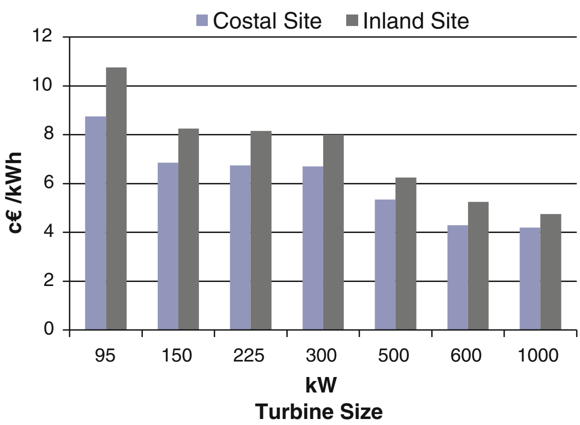
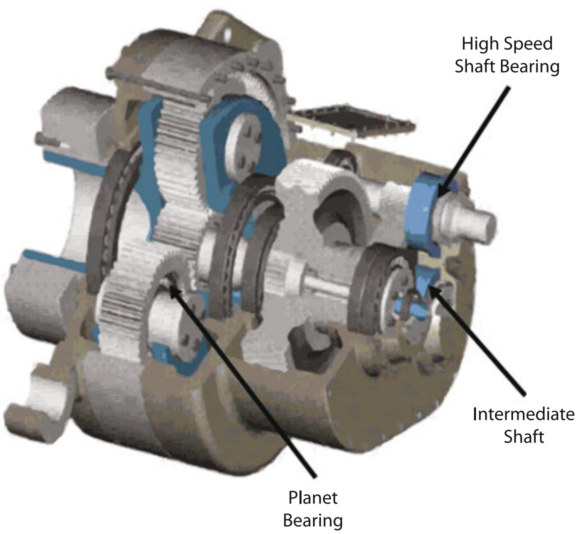


Fig. 2.2 Overall cost of generation for different turbine sizes [30]

Various types of generator are used in wind turbines. The selection of electrical generator is conducted in conjunction with that of drive train technology, principally gearbox. The majority of turbines currently in operation use indirect drive where the turbine is coupled to the generator via a gearbox. The use of a gearbox allows a high speed generator to be used which is in general more cost effective in terms of manufacture [4], direct drive systems involve low-speed electrical machines which are more expensive to produce but offer greater overall system availability due to the lack of gearbox. Wind turbines which do not use a direct drive configuration make use of a gearbox. Wind turbine gearboxes are based on well-developed and mature technologies involving gears, shafts, and bearings. Due to their size, turbine gearboxes can cost in the region of $250,000 [5]. Various investigations have been carried out into the failure rates and criticality of the various gearbox components (Fig. 2.3). A major source of wind turbine downtime is the gearbox [6–9]. Developing a cost-managed strategy to mitigate gearbox failures is difficult due to variable failure rates and a lack of accepted benchmarks [10, 11].

Gearboxes are considered to be a mature technology therefore improving design reliability is viewed as difficult. Thus detecting faults before they lead to a failure and taking steps to mitigate them is seen as an important topic. Despite their technological maturity, gearboxes continue to be seen as a major issue due to the costs involved in repairs and the reduced lifespan of 7–11 years compared to a design life of 20 years for the turbine as a whole [6]. The gearbox is one of the most expensive components of the wind turbine system and the higher than expected failure rates are increasing the cost of wind energy. To help reduce the cost of wind energy a new approach to gearbox and lubricant needs to be developed.

Fig. 2.3 Typical gearbox components



2.3 A New RCM-Based Approach to e-Maintenance of Turbines

Reliability Centred Maintenance (RCM) is a method for determining the maintenance needs of a particular asset, component, and sub-component. The focus of RCM is on maintaining system function rather than restoring equipment to an ideal condition. RCM allows the maintenance 'team' to identify the critical components, their function and failure modes, and therefore identify failure patterns and possible maintenance tasks to maintain operation. RCM, according to [12] is characterized by the following features:

- Greater safety and environmental protection, due to improved maintenance
- Improved operating performance, due to more emphasis on the maintenance requirements of critical components
- Greater maintenance cost-effectiveness, due to less unnecessary maintenance
- A comprehensive maintenance database, which reduces the effects of staff turnover with its attendant loss of experience and expertise.

Condition monitoring techniques will allow each essential/critical component, identified using an RCM analysis, to be monitored for a change in condition. Continuous monitoring using sensors, often wireless, will allow for early detection of abnormalities and corrective tasks introduced to the maintenance strategy. Condition monitoring has a number of techniques to monitor the condition of a particular component; however, the following section will focus on oil analysis. Oil degradation in wind turbine gearboxes is not detected until the gearbox has failed. Oil analysis examines the ingress of dust, water, soot, and other particles which may affect gearbox operation. A system which could measure the chemical composition of oil, continuously, and send data via wireless sensors to central 'server' would help reduce unplanned stoppages which could be extensive, costly, and in severe cases have an impact on the ecology.

2.4 Application of Oil Analysis

Oil analysis is an established and proven maintenance technique [13–15]. However, the process of performing oil analysis, involving the manual taking of samples and sending them back to a laboratory for analysis is problematic in several respects. First, the time taken for results to be obtained is problematic. Second, there is a risk of contamination due to the harsh environment of the marine environment. Thus automated, remote monitoring, and analysis have become an area of interest in the marine and offshore wind area [16, 17].

Oil analysis involves a variety of measurements which can detect a wide range of fault modes. The principal application in an epicyclic gearbox will be in the detection of wear particles. In addition to providing an early warning of developing failure, automated measurement of wear particle concentration in oil means that the lubricant can be replaced before further damage is caused by wear particles. In recent years, a variety of systems have been developed for counting and characterizing wear particles optically (c.f. [18]). Analysis in this area provides a wealth of raw data. In order for maximum benefit to be achieved it is necessary, however, to condense and interpret the data to provide operatives with appropriate diagnostic information and guidance on corrective actions necessary. Data fusion and artificial intelligence techniques can be applied to ensure that not only are the oil properties measured and displayed but faults which are indicated by these properties are identified and the appropriate maintenance can be planned and scheduled.

The remote nature of wind turbine installations means that data transmission is a key issue. Standards and protocols need to be selected which ensure that the data is securely and reliably transmitted and can be verified on reception. Furthermore, maintenance personnel need to have access to data and documentation when working in remote environments where data connectivity may be limited. It is necessary to engage with industry experts and maintenance personnel to ensure the design of data presentation is optimal and tailored to the requirements of the industry. A number of questions remain of an e-maintenance-based condition monitoring solution is to become a viable reality. These include:

- Which oil condition parameters allow detection and diagnosis of likely failure modes?
- How frequently must data be transmitted?
- How can condition monitoring limits be managed to optimize maintenance patterns?
- How can data security and integrity be managed?
- How can data be best presented to the appropriate personnel?

In order to investigate these and other important questions a test setup has been devised. This setup will build on the work of the Progressive Oil Sensor System for Extended Identification ON-Line (Posseidon) [16], [19] project which sought to address the flaws inherent in current oil analysis practice by developing and online oil condition monitoring system which will allow oil to be analyzed while still in service. The developed sensor unit is illustrated in Fig. 2.4.

It is envisaged that data display will be improved by use of a more robust unit than the laptop shown in Fig. 2.4. A professional industrial display, such as that shown in Fig. 2.5, featuring a minimum of controls and an intelligently laid out user interfaces designed in conjunction with industry would provide a suitable means for presenting data.

In order to ensure representative usage is applied to the gearbox under test, it will be turned by a motor with a power spectrum which represents the speed of the wind in a realistic fashion such as the Van der Hoven spectrum as used by Gorritxategi et al. [20, 21] (Fig. 2.6).

The load placed on the output of the gearbox is dependent on the generator and the characteristics of the electrical grid to which it is connected. This can be

Fig. 2.4 Posseidon sensor unit

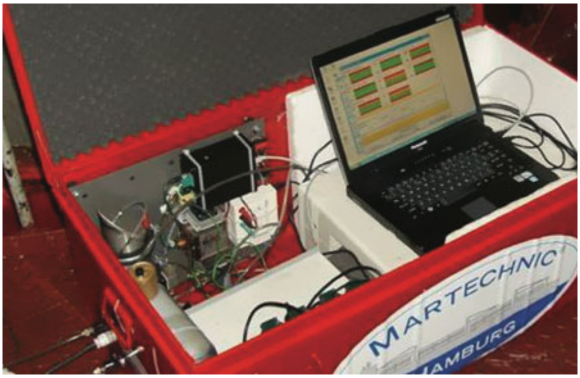
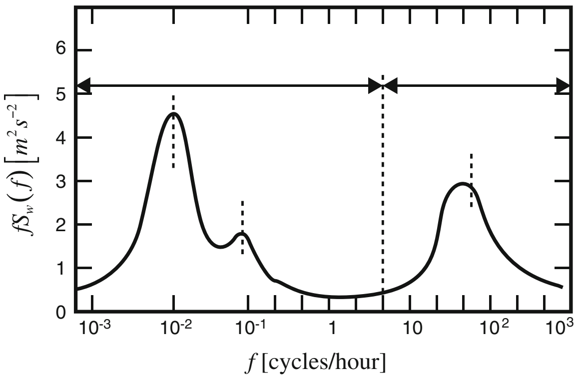


Fig. 2.5 Industrial display unit

Fig. 2.6 Van der Hoven power spectrum of wind speed



modeled and has formed the basis of test systems [22]. In order to supply such a load to the gearbox under test, a dynamometer can be used as illustrated in Fig. 2.7.

Fig. 2.7 Dynamometer for load testing gearbox

Once data has been collected it will be used to inform the development of an e-maintenance decision-making system. Data gathered will be combined with cost data and productivity measures such as availability and uptime and used to support decision making based on the RCM framework described previously. It is envisaged that an artificial intelligence tool such as Reinforcement Learning [23–25] or self organizing maps [26] will be used to optimize maintenance on a long-term basis by providing a prognosis of the condition of the turbine gearbox based on previously learned experience. Decisions can then be made on the basis of the costs of taking action versus the cost of failure and the resultant downtime. In this way, the health of a turbine can be managed in a cost-effective manner [27, 28]. In this way, maintenance can be managed according to deteriorating heath rather than through reacting to failures [29].

2.5 Conclusions

Wind turbine maintenance is predominantly a combination of time based tasks, which are undertaken at predetermined and regular-intervals, or failure-based tasks which can be costly and impact on other components. Rarely is an analysis undertaken to determine if either maintenance strategy or their tasks are appropriate, cost-effective, and necessary. It is important to develop a complete maintenance strategy utilizing the so-called "modern maintenance practices". Reliability Centered Maintenance (RCM) is a technique used mostly to select appropriate maintenance strategies for physical assets, to identify possible failure modes, causes and effects on system operation. e-condition-Based Maintenance activities are then identified and an assessment is undertaken to identify application, appropriate task selection to compliment the technology, and to maximize the return on investment.

This paper has introduced an e-maintenance Solution for Wind Turbine Gearboxes has been presented covering all aspects of development from initial conception using RCM to technological development and testing strategies. The solution will be developed into a series of 'case study tests' to determine the optimal maintenance strategy to minimize costs and unnecessary downtime. This will be achieved by utilizing recent development in senor technology, POSSEIDON. In addition, the results will show that a new approach to monitoring the oil within a gearbox is required, one which will instantly send data to a central server in which an instant diagnosis and prognosis can be determined.

References

1. World Wind Energy Report (2009) World Wind Energy Association (WWEA). Bonn, Germany
2. Botsaris PN, Konstantinidis EI (2010) Wind turbine optimisation: a holisitic approach. In: The seventh international conference on condition monitoring and machinery failure prevention technologies, 22–24 June 2010. Ettington Chase, Stratford-upon-Avon, England, UK
3. Baglee et al (2006) Modern maintenance practices: the barriers to implementation within small/medium enterprises. In: IMA international conference on modeling in industrial maintenance and reliability. Salford, England, UK
4. Polinder H, van der Pijl FFA, de Vilder GJ, Tavner PJ (2006) Comparison of direct-drive and geared generator concepts for wind turbines. IEEE Trans Energy Convers 21(3):725–733
5. Barber S, Golbeck P (2010) The benefits of a pro-active approach using preventive and predictive maintenance tools and strategies: actual examples and case studies (internet). (Cited 2010 Dec)
6. Available from: http://www.wwindea.org/technology/ch03/en/3_4_3.html
7. Puigcorbe J, de-Beaumont A (2010) Wind turbine gearbox reliability: the impact of rotor support (internet). (Updated 2010 June 3, cited 2010 Dec)
8. Available from: http://www.renewableenergyworld.com/rea/news/article/2010/06/wind-turbine-gearbox-reliability
9. Braam H, Rademakers L (2004) Models to analyse operation and maintenance aspects of offshore wind farms, ECN Report
10. Tavner P, Bussel GV, Spinato F (2006) Machine and converter reliabilities in wind turbines. In: The 3rd IET international conference on power electronics, machines and drives, Mar 2006, pp 127–130
11. Tavner PJ, Xiang J, Spinato F (2007) Reliability analysis for wind turbines. Wind Energ 10:1–18. doi:10.1002/we.204
12. McMillan D, Ault GW (2008) Condition monitoring benefit for onshore wind turbines: sensitivity to operational parameters. IET Renew Power Gener 2(1):60–72
13. McMillan D, Ault GW (2008) Specification of reliability benchmarks for offshore wind farms. In: Proceedings of the European safety and reliability, Sept 2008, pp 22–25
14. Ben-Daya M (2000) You may need RCM to enhance TPM implementation. J Qual Maint Eng 6(2):82–85
15. Meng T, Dresel W (eds) (2007) Lubricants and lubrication. Wiley-VCH, New York
16. Barnes J, Hengeveld J, Foster S, Schasfoort T, Scheele R (2004) Oil stress investigations in Shell's medium speed laboratory engine. Paper presented at: *CIMAC Congress*, Kyoto, 2004

17. Laurence RB (1994) The effect of lubrication system and marine specific factors on diesel engine emissions (MS Thesis). Massachusetts Institute of Technology 1994. Cambridge, Massachuestts, USA
18. Knowles MJ, Baglee D (2010) Condition monitoring in an on-ship environment. In: The seventh international conference on condition monitoring and machinery failure prevention technologies (BINDT CM 2010), June 2010
19. Baldwin A, Lunt S (2010) Latest developments in online oil condition monitoring sensors. The seventh international conference on condition monitoring and machinery failure prevention technologies (BINDT CM 2010), June 2010
20. Gorritxategi E, Arnaiz A, Aranzabe E, Ciria J, Terradillos J, (2006) Indirect optical measurements for lubricant status assessment. In: Proceedings of the 19th international congress of condition monitoring and diagnostic engineering management June 2006. Luleå, Sweden, pp 367–376
21. Gorritxategi E, Arnaiz A, Belew J (2007) Marine oil monitorization by means of on-line sensors. Instrumentation viewpoint, p 6
22. Neammanee B, Sirisumrannukul S, Chtratana S (2007) Development of a wind turbine simulator for wind generator testing. Int Energy J 8:21–28
23. Nichita C, Luca D, Dakyo B, Ceanga E (2002) Large band simulator of wind speed for real time wind turbine simulators. IEEE Tran Energy Convers 17(4):523–529
24. Moore I, Ekanayake J (2010) Design and development of a hardware based wind turbine simulator. In: Proceedings of the 45th international universities power engineering conference (UPEC), 31 Aug–3 Sept 2010, pp 1–5
25. Knowles MJ, Baglee D, Wermter S (2010) Reinforcement learning for scheduling of maintenance. In: The thirtieth SGAI international conference on artificial intelligence (AI-2010), Dec 2010
26. Kaelbling LP, Littman ML, Moore AW (1996) Reinforcement learning: a survey. J Artif Intell Res 4:237–285
27. Kuremoto T, Obayashi M, Kobayashi K (2005) Nonlinear prediction by reinforcement learning. In: Proceedings of the 2005 international conference on advances in intelligent computing. Springer, Berlin, pp 1085–1094
28. Wu F, Wang T, Lee J (2010) An online adaptive condition-based maintenance method for mechanical systems. Mech Syst Signal Process 24(8):2985–2995
29. Hameed Z, Hong YS, Cho YM, Ahn SH, Song CK (2009) Condition monitoring and fault detection of wind turbines and related algorithms: a review. Renew Sustain Energy Rev 13:1–39
30. Lee J, Ni J, Djurdjanovic D, Qiu H, Liao H (2006) Intelligent prognostics tools and e-maintenance. Comput Ind: e-Maintenance Spec issue 57(6):476–489. doi:10.1016/j.compind.2006.02.014

Chapter 3
A Study of Derailment in Australia: Analysing Risk Gaps with Remote Data Monitoring

G. Chattopadhyay, D. Raman and M. R. Alam

Abstract Derailment is a business risk for rail operators. However, most of the derailments are due to a combination of several causes. This paper presents a derailment study, focusing on Australia, that categorizes a comprehensive set of factors causing derailment and current mitigation techniques. Finally, a risk gap is analysed and integrated model is proposed. Further, the opportunities to incorporate remote monitoring and data analysis technologies to ensure an eMaintenance management of derailment are discussed. A case study on eMaintenance of rail lubrication, which is highly related to derailment, is also demonstrated in this paper.

3.1 Introduction

Railway is a complex system. Safe operation of this complex system is influenced by rail, wheel, rollingstock, driver- and environment-related factors along with operational and maintenance issues. In today's ever increasing demand for higher speed and axle load adds further complexity. Any unsafe condition in isolation or in combination of these leads to complex events resulting in derailment. Derailments can result in huge costs for repair and replacement of the track, vehicle, and other infrastructure along with compensation for disruption, injuries, and death-related claims. There can also be major costs due to spillage—dangerous chemicals, oil, coal, and minerals exposed to environment and some spillages can cause fires and/or explosions resulting in catastrophic problems. Some of the costs related to derailments are as follows:

G. Chattopadhyay (✉) · D. Raman · M. R. Alam
Centre for Railway Engineering (CRE), CQ University, Rockhampton, Australia
e-mail: g.chattopadhyay@cqu.edu.au

D. Raman
e-mail: d.raman@cqu.edu.au

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_3,

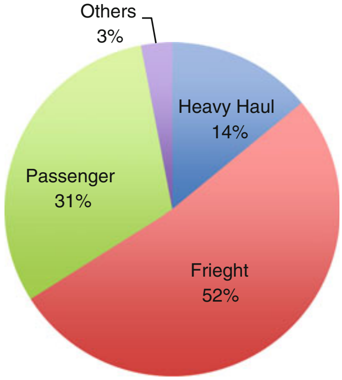
1. Direct costs of the derailment of Cairns Tilt Train, 15 November 2004 are estimated at AUD $35.5 million [1].
2. A massive Rio Tinto iron ore train derailed on 30 January 2009, spilling thousands of tonnes of iron ore. Eighty wagons came off the rails near a junction point, 80 km north of Mt Tom Price. The damage bill is estimated at up to AUD $10 million and downtime of 5 days and loss of 800 m of track [2].
3. In 2000, the Hatfield accident in the UK killed 4 people and injured 34 others and has lead to a cost over £733 million for repairs and compensations [3].
4. In 1977, the Granville train disaster in NSW, Australia killed 83 people and injured 213 others [4].

There are huge cost in terms of business loss reduced and exports in addition to these direct costs. Hence, this paper provides a study on derailments in Australia, the various factors related to the derailments and risk gaps in terms of evaluation for a comprehensive analysis.

3.2 Derailments: An Overview

The first step toward studying of derailments is identifying the key factors that cause derailment. Of the major derailments and accidents between 2001 and 2009 in Australia 73 are analyzed here. The modes of operation for these derailments/accidents are shown in Fig. 3.1. The main causes are summarized in Fig. 3.2. A detailed list of causes are categorized and represented in Fig. 3.3 and potential losses are categorized in Fig. 3.4. From the investigation analysis, it is observed that track related issues and issues related to safety practices are the major influencing factors. Track related causes include: rail break due to rolling contact fatigue (RCF); track misalignment; rail buckling and rail geometry issues; track gauge widening; and many other factors. The stability of the track is influenced by changes in rail temperature.

Fig. 3.1 Derailment in Australian rail operations



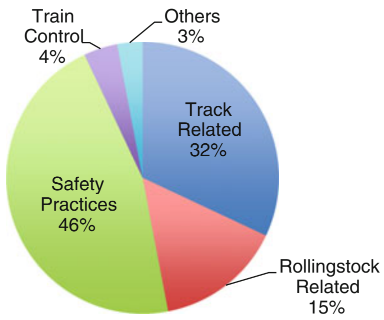


Fig. 3.2 Derailment causes in Australia

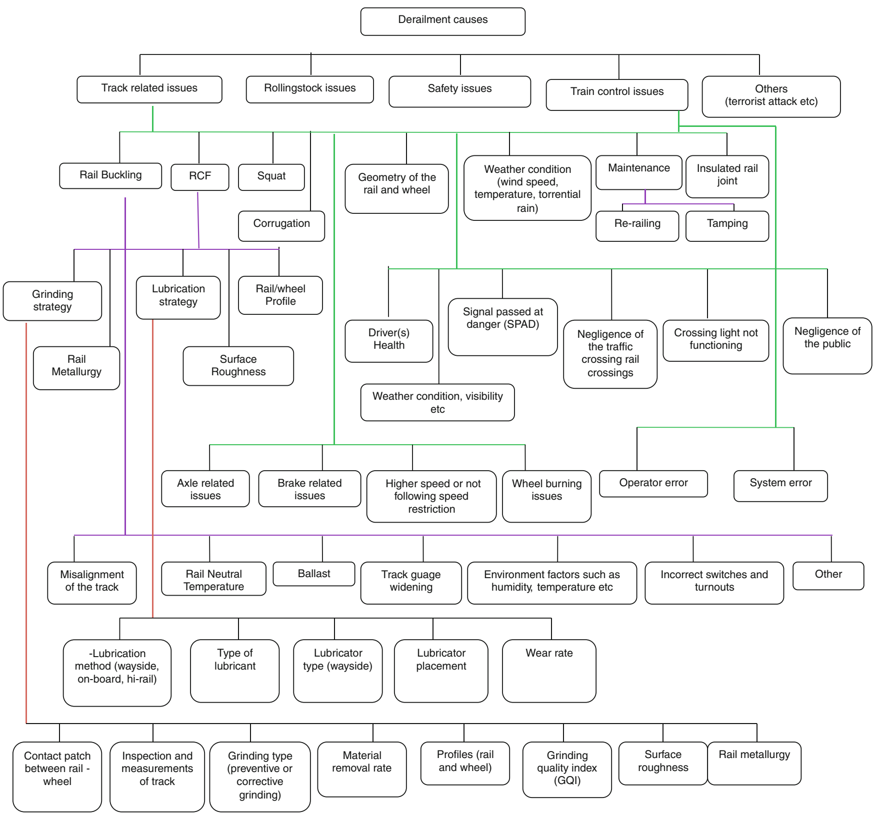


Fig. 3.3 Categorization of derailment causes

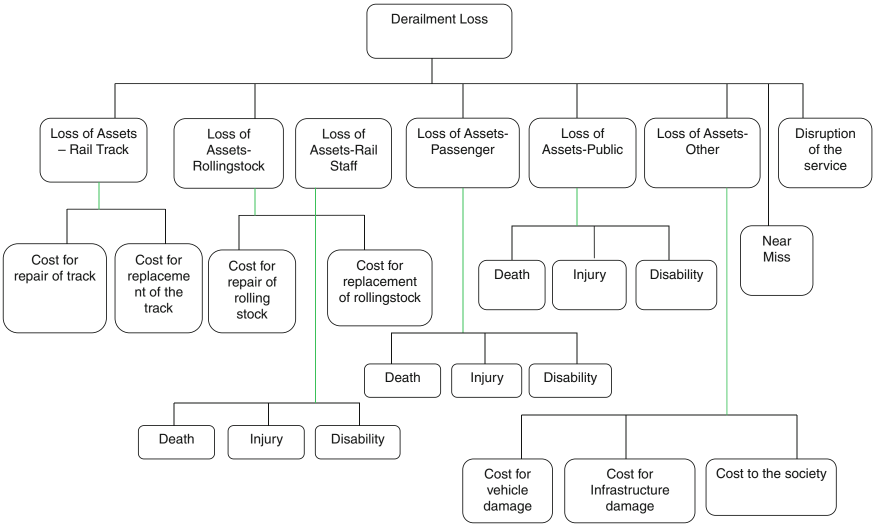


Fig. 3.4 Potential losses due to derailment

Rail Breaks: In the past, it was common for a rail break to start near the joint between discrete rail segments. Manufacturing defects in rails can also cause rail breaks. Wheel burns contribute to rail breaks by changing the metallurgy of a rail. Rails are more likely to break when the weather is cold, when the ballast and ties/sleepers are not providing as much support as they should, and when ground or drainage condition is such that 'pumping' occurs under heavy load.

Rolling contact fatigue (*RCF*): Each year in Europe, there is several hundred broken rails due to cracks initiated by RCF [3].

Track Buckling: The amount of rail expansion depends on the rise in temperature, and the coefficient of thermal expansion. If left unchecked, that expansion would cause the track to buckle. In 2006, there were 50 track buckling related derailments with $13 million reportable damages in USA, and in 2007, the reportable damages rose to $14 million with 34 track buckling-related derailments [5].

Continuous welded rail (*CWR*) is replacing jointed track for the advantages of better economics of maintenance and enhanced ride comfort. A well-known risk with CWR, however, is its potential for buckling due to high thermally induced compressive loads.

Track misalignment: It is predominantly associated with track buckling. In addition, misalignments occur under thermal load and vehicle passage which can cause the track to suddenly snap into a buckled configuration [6].

Weather condition: Environmental factors such wind speed and rain have adverse effects on rail stability. The National Safety Transportation Board of USA determines that the probable cause of the April 18 2002, derailment of Amtrak Auto Train, where four passengers were killed, was a heat-induced track buckle that developed because of inadequate track-surfacing operations, misalignment of curve, insufficient track restraint, and failure to re-establish an appropriate neutral rail temperature [7].

Corrugations: Its formation and development causes fierce vibrations of structures of the railway vehicle and track, as well as noise, and also reduces the life of structural parts and can potentially lead to derailment [8].

Fatigue: In recent years, higher train speeds and increased axle loads have led to larger wheel/rail contact forces. Unlike the slow deterioration process of wear, fatigue causes abrupt rail and wheel breaks. These failures may cause damage to rails, damage train suspensions, and derailment of the train [9].

Excessive speed also causes derailments.

1. Wheel climb, in which the wheel is lifted off the track because the friction problem between the flange and the gauge face, causing the wheel flange to climb outward over the rail head.
2. Rail roll, in which the horizontal force applied by the flange to the gauge face of the rail is high and leads to overcoming the anchoring forces of rail spikes and clips.

3.3 Current Practices to Prevent Derailment

Rail maintenance takes up a very large portion of a rail operator's budget. The Transportation Research Board reported that, in 2004, a little more then US$10 billion was spent in USA on rail transit system maintenance [10]. The industry practice rail grinding, rail lubrication (Eurostar saved £ 1 million a year on maintenance and wheel replacement by improving lubrication [11]), inspections, and replacements in order to mitigate the derailment issues. They maintain tracks to mitigate the buckling, problems in welding, and repair to maintain rail neutral temperature. Many rail operators issue speed reduction orders when daily ambient temperatures reach a certain level in summer [12]. However, excessive speed restriction costs the rail industry millions of dollars each year. Rails employ track stabilization following a ballast disturbance such as surfacing, realignment, or ballast renewal [13].

Rail operators also use rail anchors to limit the longitudinal expansion of the rail. Many rail operators are replacing timber with concrete sleepers with high tonnage or high-speed trains in order to reduce out-of-gauge derailments. Zhao et al. [14] proposed a model to analyze the risk of derailment by using a probabilistic approach to model the development of rail defects leading to rail breaks and derailment. The risk of derailment is measured by the expected number of rail breaks multiplied by the severity of rail break. Bagheri et al. [15] proposed improved operational strategies with a specific focus on dangerous goods marshalling practice in the train assembly process for mitigating derailment risk. The computationally intractability of current models have led to heuristic method based on genetic algorithm. Appraisal of track degradation related to vehicle and track is carried out using "DECOTRACK" [16]. Jin et al. [8] simulated the formation and evolution of corrugation on a curved rail, combining the three dimensional rolling contact model of non-Hertzian form. VAMPIRE^{TM} is used to analyze the forces on the vehicle. Rail Sciences Inc. (RSI), USA, has investigated over 1,500 individual derailments for many domestic and international railroads. Chattopadhyay [3, 4] proposed an integrated model and developed a prototype simulation software for grinding intervals, inspections intervals, lubrication, and replacement decisions to reduce risk and cost in rail operations.

3.4 Risk Gaps Analyzed in Derailment Study

A comprehensive framework (Fig. 3.5) is proposed here. Chattopadhyay et al. [17] have developed model to select optimal rail lubricants especially for heavy haul lines in Australia.

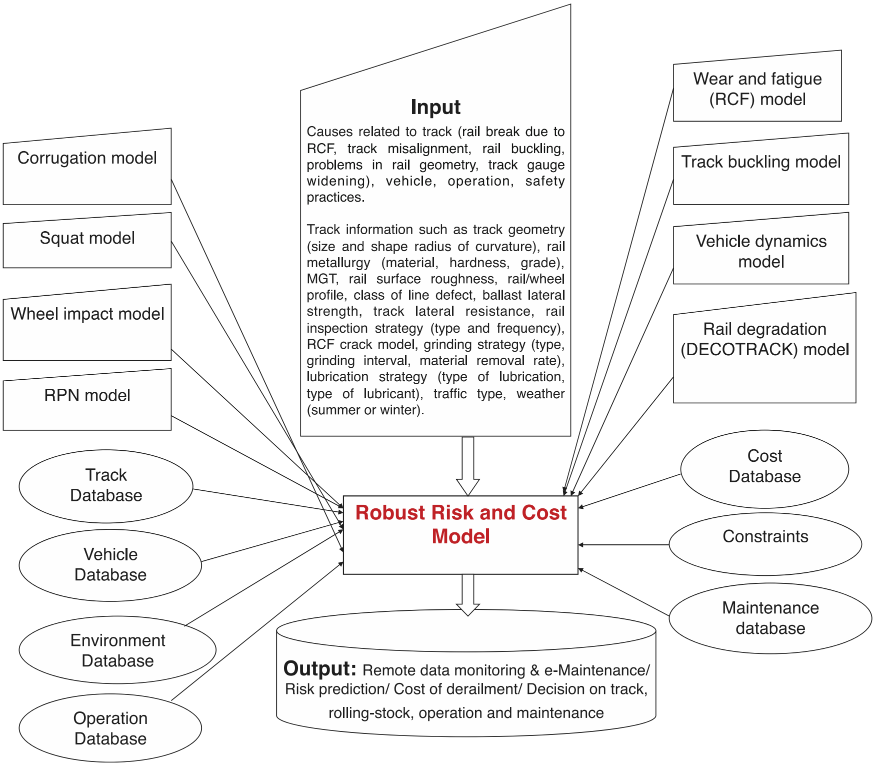


Fig. 3.5 Integrated model for risk and cost analysis of derailment

The selection of lubricant needed to predominantly consider the traffic and weather conditions in Australian network. Similarly, they have proposed model to apply right quantity of lubricant at right place and time [18]. Chattopadhyay et al. [3, 4, 19] have developed model to determine the grinding cycle with respect to MGT (Million Gross Tonnes) of load. Integrated risk and cost analysis model is proposed in this paper to estimate the probability of derailment at various segments of rail network for developing appropriate preventive maintenance strategy.

3.4.1 Remote Data Monitoring and Analysis of Derailment Risks

As discussed above, predicting the derailment in a rail network involves a complex variety of factors. Consequently, managing this derailments is a difficult task and consumes a significant amount of labor hours and resources. Whereas, employing remote monitoring technologies to access the track, rolling stock, and environmental condition data from the network can help reduce uncertainties in remote locations, as well as the need for staff to attend the sites in person and work in trackside danger zones.

For a large rail operator, traditional maintenance can be a complex and costly operation. With changing traffic volumes it can be even more difficult to schedule track maintenance. Just accessing the tracks for maintenance can be troublesome, especially in high traffic corridors. Many rail operators work in a zero harm environment and have very rigid track access requirements that need to be met for access [20]. Accessing track in tunnels, cuttings, and where there is poor trackside access may require the maintenance staff to take possession of the track—halting trains for inspections and maintenance—which has a direct impact on train operations. As an alternative, real-time monitoring and tracking of data give track maintainers a greater understanding of track conditions, risk level of derailments, and thus reduces the cost of managing the tracks and risks of derailments. Such remote monitoring technologies are being slowly being introduced in Australian rail network. This low rate of adoption is due to the challenges in providing capital budget and support at remote locations.

Such a remote monitoring setup consists of a data logger that is connected to a wireless transmitter or cable. There are many ways to transmit data using wireless technology. These include GPRS (General packet radio service) (cellular technology), RF (radio frequency), and Microwave. The set-up cost involved with microwave is very expensive, is mostly suitable over line-of-sight transmission and signal quality gets downgraded by rain, dust, snow, and fog. Data transmission over railway's own copper wire-based signaling network is also possible. Whereas, railway operators often prefer a dedicated signaling network and therefore this is often not used [20].

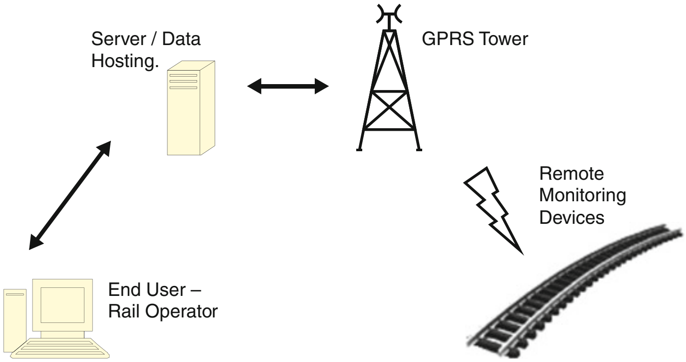


Fig. 3.6 Remote condition monitoring for derailment

RF signals have a limited range and requires repeaters over long transmission distance. GPRS technology provides data transmission at higher speeds and longer distances. The selection of GPRS is dependent on cellular network coverage that provides data transmission capability. This includes the 2G and 3G service. The 3G service can utilize the latest HSDPA (High-Speed Downlink Packet Access) for faster data transmission compared to the 2G service which is much slower. With 3G data transfer rates of up to 10.5 Mbits/s can be achieved. Also available is an option for updates to mobile phone via an SMS or a system-generated e-mail to a designated email account.

Figure 3.6 illustrates the flow of information as to how the data would be acquired by the end user. In this system, data is transmitted via the mobile network using a GPRS tower to the Data hosting Server where the data is processed and analysed. The end users log into the data hosting server to retrieve data and carry out further analysis if required. Some manufactures of detection and condition monitoring also provide real time access to data. There is also an option to directly communicate with the remote monitoring devices to change settings based on operating, track, and environmental conditions.

3.4.2 Case Study on eMaintenance: Gladstone, Queensland

Effective friction management of the wheel rail interface plays an important role in prolonging the life of rail and wheel and significantly controls the risks of derailment. An electronic wayside lubricator with remote performance monitoring was setup for this trial and on a track on a heavy haul coal line north of Gladstone.

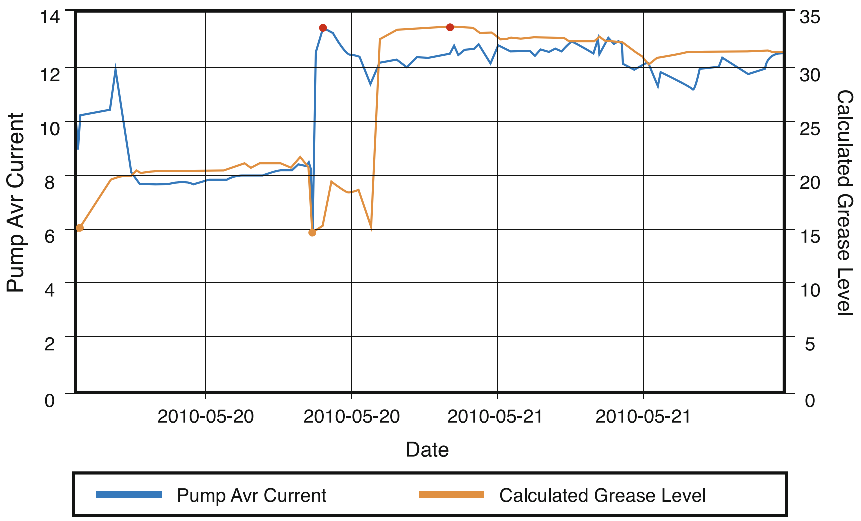


Fig. 3.7 Drop in average pump current due to an air pocket with no change in product level. Portec Rail Products Inc, www.portecrailrpm.com

After commissioning of the lubricator it was tested to check all the functionality which it passed. A rapid drop in pump motor current was recorded during the trial (Fig. 3.7). This was due to an air pocket and resulted in preventing the grease being pumped. This resulted in the rails running dry, which can in some cases increase the possibilities of derailment. Another opposite case would be too much lubricant on the track and contamination .

3.4.3 *Cost Benefits of Adopting Remote Monitoring*

Rail operators would need to decide on eMaintenance and better way to justify the technology is to perform a cost–benefit analysis. Consequently, a cost analysis was performed for the above case study. It was established the financial benefits of eMaintenance technology in addition to the operational benefits outweighs the investment cost and the uncertainly of the condition in between inspections. Although the investment is initially higher compared to current practice, it has a much lower operational cost and there are a reduced number of onsite call and reduced risk of in service failure and undetected problems. These encouraging results provide enough support to extend the technology for managing the derailments across the longer rail networks in Australia.

3.5 Conclusion

Railway is of strategic importance for Australia. A majority of the coal mines link to the ports through railway. Huge amounts of freight are moved all over the Australia via rail networks. Unreliability of rail network affects the supply chain resulting in huge economic loss, particularly in the mining and resources sector. Derailment reports are analyzed for main causes for derailment in Australia. Current models/practices are discussed and they are found to be inadequate to predict derailments accurately and are limited in risk and cost estimation. Thus a robust model is proposed along with supporting case of using eMaintenance and data analysis for predicting and managing complex events similar to derailments. The authors of this paper have developed already key individual modules for the proposed robust model.

References

1. Australia (2004) Derailment of cainrs tilt train VCQ5. Australian Transport Safety Bureau, North of Berajondo, Queensland, 15 Nov 2004. Report No. 2004007
2. 80 carriages derailed in $10 m WA train accident. ABC News Australia (Internet), 30 Jan 2009 (cited 2011 Mar 16). Available from: http://www.abc.net.au/news/stories/2009/01/30/2478649.htm
3. Chattopadhyay G, Reddy V, Larsoon-Karik PO (2005) Decision on economical rail grinding interval for controlling rolling contact fatigue. Int Trans Oper Res 12(6):545–558
4. Chattopadhyay G (2009) Decision support system for integrated rail maintenance. In: AusRail conference 2009, Adelaide, Australia
5. Samavedam G, Kidd J (2008) Development of rail neutral temperature monitoring device. US Department of Transportation, Washington, DC, USA, Report No.: RR 08-31, Dec 2008
6. Sussmann T, Kish A, Trosino M (2003) Investigation of the influence of track maintenance on the lateral resistance of concrete tie track. In: Proceedings of transportation research board conference 2003, Washington, DC, USA. Cambridge. John A. Volpe National Transportations Systems Center, 2007, MA, USA
7. Drago JJ (2008) Train car compensator for platform gap spacing. US7908976, 2 July 2008
8. Jin X et al (2005) Numerical simulation of rail corrugation on a curved track. Comput Struct 83(25–26):2052–2065
9. Liu Y, Stratman B, Mahadevan S (2006) Fatique crack initiation life prediction of railroad wheels. Int J Fatique 28(7):747–756
10. Daniels LE (2008) Track maintenance costs on rail transit properties. Transport Research Board, Washington, DC, USA, Sept 2008
11. Larke R (2003) Survey of rail/wheel lubrication practice. UK Rail Safety and Standards Board, London, England, 7 Mar 2003. Report No.: ITRL—T12442-001
12. Kish A, Clark DW (2009) Track buckling derailment prevention through risk-based train speed reductions. In: AREMA annual conference 20–23 Sept 2009, Chicago, Illinois, USA
13. Krish A, Sussmann T, Trosion M (2003) Effects of maintenance operations on track buckling potential. In: Proceedings of international heavy haul association, May 2003, Fort Worth, Texas, USA
14. Zhao J, Chan AHC, Burrow MPN (1995) Probablistic model for predicting rail breaks and controlling risk of derailment. J Transp Res Board 1995:76–83 23 Oct 2007

15. Bagheri M, Saccinabbi FF, Fu L (2009) Risk-based model for marshalling dangerous goods railway cars in rail yards. In: Transportation research board 88th annual meeting, 11–15 Jan 2009, Washington, DC, USA
16. Larsson D (2002) Verification of the DeCoTrack: a prediction model for degradation cost of track. Lulea University of Technology, Lulea
17. Desai A et al. (2011) Experimental setup and methodology for the analysis of rail lubricant effectiveness. In: Proceedings of the 24th international congress on condition monitoring and diagnostics engineering management, May 30–Jun 1, 2011. COMADEM international, Stavanger, Norway, 2011
18. Uddin MG et al. (2010) Wayside lubricator placement model for heavy haul lines in Australia. In: CORE 2010: rail, rejuvenation and renaissance, Wellington, NZ, 14 Sept 2010. Engineers Australia, Wellington, NZ, 2010
19. Reddy V et al (2007) Modelling and analysis of rail maintenance cost. Int J Prod Econ 105(2):475–482
20. Desai A, Chattopadhyay G, Howie A (2010) Remote condition monitoring and control of wayside rail lubricators. In: The 1st international workshop and congress on e-maintenance, Lulea, Sweden, 22–24 Jun 2010. Division of Operation and Maintenance Engineering, Lulea University of Technology, Sweden 2010

Chapter 4
eMaintenance Industrial Applications: Issues and Challenges

R. Karim, A. Parida, O. Candell and U. Kumar

Abstract Achieving business excellence within industries which utilize complex technical systems requires effective and efficient maintenance. Maintenance is an important enabler of business performance. An effective maintenance strategy creates additional values in an organization's value-generating process. Establishing an effective and efficient maintenance process is highly dependent on supporting Information and Communication Technology (ICT) infrastructure for information logistics that facilitates maintenance decision support. eMaintenance solutions facilitate effective and efficient management and control of maintenance activities through an enhanced utilization of computing; like estimation of Remaining Useful Life (RUL); reduction of No-Fault Found (NFF); and prediction of fault. eMaintenance solutions enable a seamless integration and fusion of information services provided by information intensive systems with embedded components in order to manage increasing and extensively distributed real-time data collected via different data sources (e.g., sensors). Since, the emerging 'Internet of Things' is expecting to dramatically change information systems with inherent embedded components, eMaintenance solutions need to be adapted to this new context to fulfill the overall business requirements on an effective and efficient decision-making process; for e.g.,: real-time analysis based on real-time data and context-aware information logistics. However, development and establishment of

R. Karim (✉) · A. Parida · U. Kumar
Division of Operation and Maintenance Engineering, Luleå University of Technology, 971 87 Luleå, Sweden
e-mail: ramin.karim@ltu.se

A. Parida
e-mail: aditya.parida@ltu.se

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

O. Candell
Saab Support and Services, Linköping, Sweden
e-mail: olov.candell@saabgroup.com

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_4,

proper eMaintenance solutions can be facilitated through utilization of an appropriate framework which deals with real-world challenges. This paper explores some of the issues and challenges pertaining to eMaintenance in industrial applications.

4.1 Introduction

Business sustainability, coupled with advancements in technology, under a global competitive and economic recession scenario, have compelled the manufacturing industry and services to focus on e-health monitoring of their engineering asset for the enhanced value and improvement in asset management. The prognostic business needs compels organizations to minimize the manufacturing and service downtime by reducing the asset performance degradation. These requirements necessitated the development of proactive maintenance strategies to provide optimized and continuously improved performance with minimized system breakdown and maintenance. It is in this context that the eMaintenance solutions will contribute by providing technologies and methodologies to retrieve useful information. Maintenance can then be carried out in a more efficient way, leading to improved Overall Equipment Effectiveness (OEE), sustainability and improved service business. A more up-to-date term to describe a new approach is “Business Centered Operational Decisions”, i.e., the approach that takes into account all the aspects of modern business domain. Regardless of the maintenance strategy, it can be noted that information and information provisioning are essential in order to support an effective and efficient decision-making process related to operation and maintenance in an organization.

Today, information needs and the development of Information and Communication Technology (ICT) have added velocity to everything that is done within industry through transforming the business process into eBusiness [1]. In the context of maintenance related to complex technical systems, information logistics addresses challenges to provide eMaintenance solutions that deal with: (1) time management, which addresses ‘when to deliver’; (2) content management, which refers to ‘what to deliver’; (3) communication management, which refers to ‘how to deliver’; and (4) context management, which addresses ‘where and why to deliver’. A proper eMaintenance solution enables context-adapted and just-in-time access to maintenance information through information services. For information logistics, the right information in the right formats is required to be delivered at the right time and place to the right person making the decision [2, 3].

There are extensive existing efforts that can be adapted to approach issues related to some of these needs. These efforts can be categorized in two groups: (1) generic efforts that are not specifically aimed at maintenance, but might be utilized for maintenance purposes and (2) specific efforts that are specifically aimed at eMaintenance. Examples of more maintenance-specific efforts are: Product Life Cycle Support (PLCS), Production Planning and Scheduling (PPS), ISA-95, Open System Architecture for Enterprise Application Integration (OSA-EAI), Common Relational Information Schema (CRIS), Condition monitoring and diagnostics of machines—Data processing, communication, and presentation (ISO 13374), the PROTEUS-platform, the concept of System for Mobile Maintenance (the SMMART-project) with focus on smart tags and wireless communication, TeleMaintenance platform (TELMA), Technologies and Techniques for new Maintenance concepts (within the TATEM-project) with focus on monitoring technologies related to aviation, Maintenix (MXI Technologies), IFS Applications and the SAP Service and Asset Management within the SAP Business suite. Some of these efforts contemplate propose support solutions from a business process perspective (e.g., RUP and EssUP), while others focus on technological aspects of service provision (e.g., OSA-EAI, PROTEUS, DYNAMITE, and Maintenix). To enhance the information logistics related to maintenance support through appropriate information services, both the business process and related technologies need to be linked and aligned seamlessly (see the reasoning in [3–6, 7].

Examples of generic efforts are: Essential Unified Process (EssUp), Rational Unified Process (RUP), Service-Oriented Architecture (SOA), Event-Driven Architecture (EDA), Web services, Extensible Markup Language, XMLSchema, Reference Model for Service Oriented Architecture, Open Document Format, Microsoft.NET and Biztalk, IBM Web Sphere, Oracle BEA WebLogic, and SAP Business suite.

There are several existing contributions that deal with the establishment of eMaintenance from a software (SW) architecture perspective. One such contribution is the concept of a common database, i.e., an integration at the data level for provision of the needed flexibility for different working areas (e.g., maintenance), as support to strategic, tactical, and operational decision making [7]. Another contribution is based on a client–server approach [8]. This client–server architecture is based on two subsystems: (1) a maintenance center, which is intended to be tailored to the needs of the company through cooperation among the maintenance-related areas; and (2) local maintenance, which implements and evaluates the maintenance system provided by the maintenance center. Another architecture approach relies on a star-like layered architecture [9, 10]. One way to realize this star-layered architecture is to use a platform that utilizes the Web services technology upon the existing maintenance applications to provide function and data integration. However, the Web services technology is only one way to realize a Service-Oriented Architecture (SOA). An additional contribution is a proposed eMaintenance Management Framework (eMMF), which is based on a service-oriented approach for the development of eMaintenance solutions [11].

When considering SW-related architecture, it can be concluded that SW architecture has evolved from focused on server-computing to client-computing and during the last decades the focus has been changed to distributed computing. Distributed computing has been manifested through various approaches, e.g., client–server, object oriented, component based, and service oriented. The latest architectural approach is sometimes called “virtual computing”, which provides

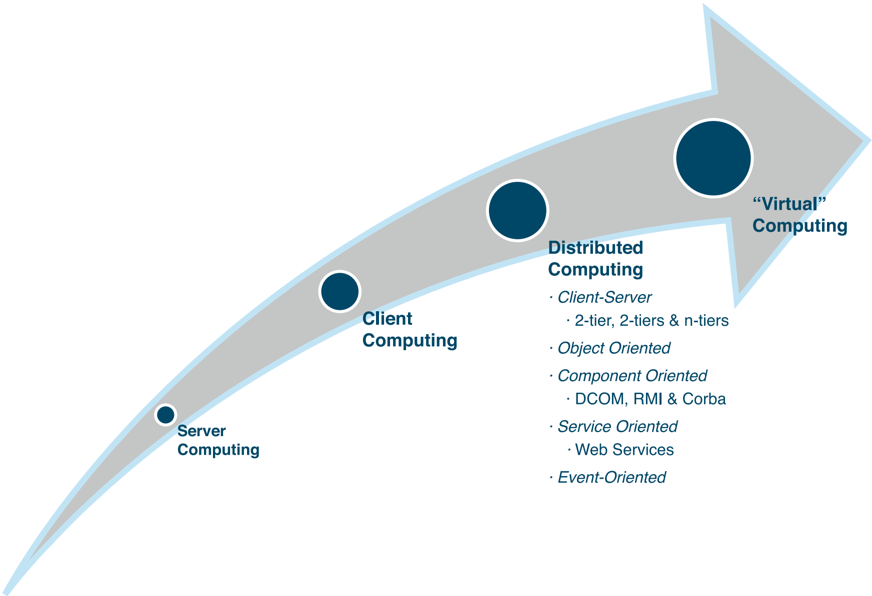


Fig. 4.1 The architecture maturity [12]

an abstraction level of distributed computer resources, e.g., disk spaces, SW applications, and processor capacity. Examples of providers that offer support for virtual computing are: HP, Sun, IBM, and Microsoft. See Fig. 4.1.

The latter contribution provided by Jantunen and Prakash [13] is described in more detail in following section.

4.2 eMaintenance Solutions Challenges

Today, eMaintenance has reached a high degree of attention within the industry [13–20]. eMaintenance is considered as an enabler to achieve business excellence within industries which utilize complex technical systems. It facilitates establishment of an effective decision-making process related to maintenance by providing proper tools, technologies and methodologies based on enhanced utilization of computing and Information and Communication Technology (ICT), see Fig. 4.2. Services provided by eMaintenance solutions can be utilized to, e.g., estimate Remaining Useful Life (RUL) of assets; reduce appearance of the No-Fault Found (NFF) phenomenon; assess performance; monitor health and usage of assets; and predict fault through.

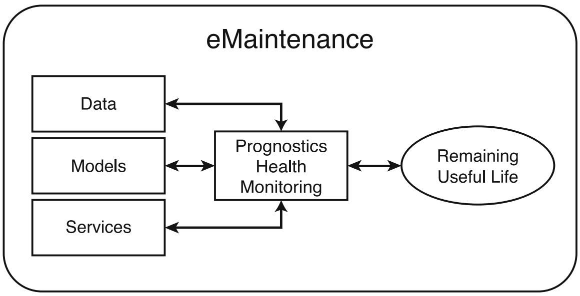
Development of eMaintenance solutions can be utilized by industry to achieve business excellence requires corresponding tools, methodologies, and technologies

Fig. 4.2 eMaintenance for decision support

to facilitate maintenance decision-making process through establishment infrastructure for implementing Condition Based Maintenance (CBM) strategies and Predictive Health Monitoring (PHM). In this context, fusion of data (e.g., business data, product data, maintenance data, condition monitoring data, and operational data) and services that process the data is essential, as shown in Fig. 4.3.

Fusion of this extensive set of data and services requires overarching strategies the enable development of SW and HW from a holistic system perspective. One approach that can be utilized in this context is Service-Oriented Architecture (SOA). The Organization for Advanced Standards for Information Society (OASIS) describes SOA as a collection of best practises, principles, and patterns related to service-aware, enterprise-level, and distributed computing [21]. SOA-related standardization efforts at OASIS focus on workflows, translation coordination, orchestration, collaboration, loose coupling, business process modeling, and other concept that support agile computing [21]. Hence, service-orientation might be considered as an architectural design philosophy that facilitates encapsulation of business logic in services. Hence, SOA-based eMaintenance solutions are considered as a set of services aimed to provide capability for maintenance decision support. These services are able to gather and process data related to the system-of-interest. Since, SOA enable development of loosely coupled and autonomous services, SOA-based eMaintenance solutions can be distributed at different levels and embedded into the system components, as illustrated in Fig. 4.4. Embedded eMaintenance services aims to improve maintenance

Fig. 4.3 Fusion of eMaintenance for decision support



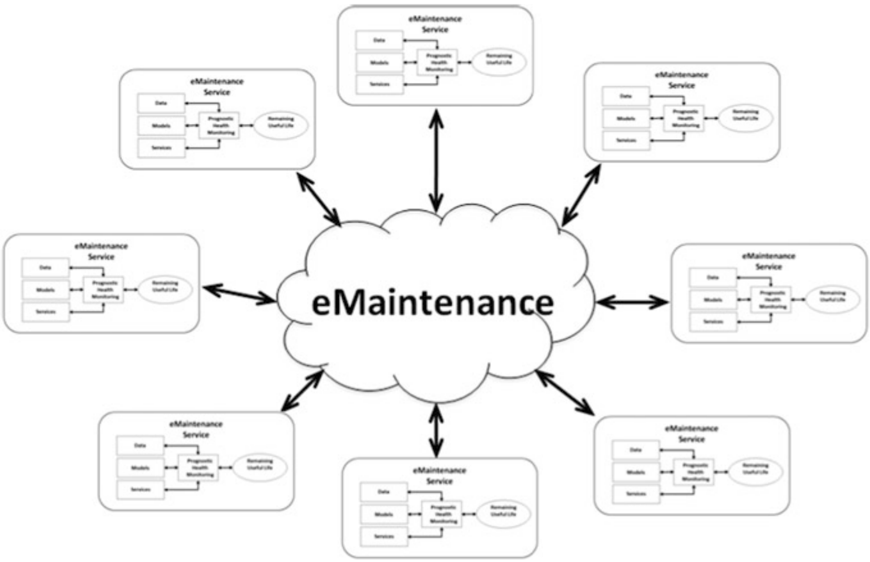


Fig. 4.4 Concept of SOA-based eMaintenance solution

information logistics and increase the effectiveness and efficiency of data fusion from various data sources related to maintenance.

However, development of eMaintenance for industrial application faces a number of challenges which can be categorized into: (1) Organizational; (2) Architectural; (3) Infrastructural; (4) content and contextual; and (5) integration.

Organizational challenges mainly focus to aspects related enterprise resource management. Examples of these challenges are: (1) restructuring of the organizations involved in eMaintenance; (2) planning of resources (e.g., material, spare-part); (3) information management; (4) knowledge management; and (5) management of heterogeneous organizations.

Architectural challenges deal with issues related to the overall architecture of eMaintenance solutions. Some of these challenges are: (1) development of a framework for development of eMaintenance; (2) development of models for decentralized data processing and analysis; (3) development of service model for decentralized data analysis; (4) development of model based prognostic tools; (5) development of model aimed for data and information visualization to support Human–Machine Interaction; and (6) development of model aimed for distributed data storage capability.

Infra-structural challenges address with issues related to provision of necessary technologies and tools that are required to meet needs and requirements when services, according to SOA, are developed, implemented, and managed in an enterprise. Example of these are: (1) network infrastructure (e.g., wired and wireless); (2) authentication of services and users; (3) authorization of services and users; (4) safety and security mechanism; (5) maintainability of eMaintenance services; (6) availability performance management; and tracing and tracking mechanism.; and (7) provision of mechanism aimed for documentation and archiving.

Content and contextual challenges are mainly related to data and information provided through the eMaintenance services. Some of these challenges are: (1) provision of appropriate ontology through which data from data sources (e.g., process data, product data, condition monitoring data, and business data) can smoothly and seamlessly be integrated; (2) provision of quality assurance mechanism that ensures that required data quality is fulfilled and visualized, in order to increase the quality of decision making; (3) mechanism for sensing user's current situation in order to adapt information to user's context; (4) provision of mechanism for describing various context; (5) mechanism to manage uncertainty in data sets; and (6) provision of mechanism for pattern recognition.

Integration challenges address issues related to coordination, orchestration, and integration of services and data managed by the eMaintenance solution. Some of these challenges are related to: (1) management, interaction, and interactivity of services; (2) configuration management of eMaintenance services; (3) enablement of integration capability across a multiplatform and technologies; and (4) reduction of appearance of data-islands.

4.3 Conclusions

This paper explores the challenges and issues pertaining to the implementation of eMaintenance in industrial applications. Since, it is not possible to specify all the common issues and challenges for all types of industries and services, some of the most relevant issues and challenges are presented. It can be concluded that development of eMaintenance solution in industry today faces some essential challenges. These challenges are related to a wide spectrum of domains and disciplines, e.g., maintenance engineering, computer science, and human factors. This means that during the development of eMaintenance solutions, industry needs to consider the challenge of managing knowledge and tools (e.g., standards, technologies, and methodologies) in this multidisciplinary domain. Furthermore, all the industry specific relevant issues, needs to be considered and addressed for achieving successful eMaintenance application facilitating effective and efficient management and control of maintenance activities through enhanced computing utilization.

References

1. Lee J (2003) E-manufacturing–fundamental, tools, and transformation. Robot Comp Integr Manuf 19(6):501–507
2. Parida A, Kumar U (2004) Managing Information is key to Maintenance Effectiveness. In: Proceedings of the intelligent maintenance systems (IMS) international conference, Arles, France, 15–17 July 2004
3. Parida A, Galar D, Berges L (2011) Decision support systems in maintenance: fusion of business data and physical data. In: The 8th international conference on condition monitoring and machinery failure prevention technologies, Cardiff, England, 20–22 June 2011
4. Liyanage JP, Kumar U (2003) Towards a value-based view on operations and maintenance performance management. J Qual Maintenance Eng 9(4):333–350
5. Söderholm P, Holmgren M, Klefsjö B (2007) A process view of maintenance and its stakeholders. J Qual Maintenance Eng 13(1):19–32
6. Koc M, Lee J (2001) A system framework for next-generation E-Maintenance Systems. In: Proceedings of the 2nd international symposium on environmentally conscious design and inverse manufacturing, Tokyo, Japan, 11–15 Dec 2001
7. Kans M (2008) On the utilisation of information technology for the management of profitable maintenance. Doctoral thesis, Växjö University, Växjö, Sweden
8. Han T, Yang B-S (2006) Development of an e-maintenance system integrating advanced techniques. J Comput Ind 57(6):569–580
9. Bangemann T, Rebeuf X, Reboul D, Schulze A, Szymanski J, Thomesse J-P, Thron M, Zerhouni N (2006) PROTEUS—creating distributed maintenance systems through an integration platform. J Comput Ind 57(6):539–551
10. Bangemann T, Reboul D, Szymanski J, Thomesse J-P, Zerhouni N (2004) PROTEUS—an Integration Platform for Distributed Maintenance Systems. In: Proceedings of the intelligent maintenance systems (IMS) 2004 international conference, Arles, France, 15–17 July 2004
11. Karim R (2008) A service-oriented approach to eMaintenance of complex technical systems, Doctoral Thesis, Luleå University of Technology. Sweden
12. Söderholm P, Karim R, Candell O, Kumar U (2009) In Proceedings of CM/MFPT 2009—(CM2009). In: CD proceedings of the sixth international conference on condition monitoring and machinery failure prevention technologies—CM/MFPT 2009, held 23-25 June 2009 at the Stillorgan Park Hotel, Dublin, Ireland. British Institute of Non-Destructive Testing and Coxmoor Publishing Campos
13. Jantunen JE, Prakash O (2009) A web and mobile device architecture for mobile e-Maintenance. Int J Adv Manuf Technol 45(1–2):71–80
14. Iung B, Crespo Marquez AA (2006) Editorial: special issue on eMaintenance. J Comput Ind 57(6):473–606
15. Kans M (2008) On the utilisation of information technology for the management of profitable maintenance. Doctoral thesis, Linnaeus University, Växjö, Sweden
16. Karim R (2008) A service-oriented approach to emaintenance of complex technical systems. Doctoral thesis, Luleå University of Technology, Luleå, Sweden
17. Lee J, Ni J, Djurdjanovic D, Qiu H, Liao H (2006) Intelligent prognostics tools and e-maintenance. J Comput Ind 57(6):476–489
18. Muller A, Marquez AC, Iung B (2008) On the concept of eMaintenance: review and current research. J Reliab Eng Sys Safety 93(8):1165–1187
19. Tsang A (2002) Strategic dimension of maintenance management. J Qual Maintenance Eng 8(1):07–39
20. Koc M, Lee J (2001) A system framework for next-generation E-Maintenance Systems. In: Proceedings of the 2nd international symposium on environmentally conscious design and inverse manufacturing 2001, Tokyo, Japan, 11–15 Dec 2001
21. Kajko-Mattsson M, Karim R, Mirjamdotter A (2011) Essential components of eMaintenance. Int J Performability Eng 7(6):505–521

Chapter 5
Aspects of Data Quality in eMaintenance: A Case Study of Process Industry in Northern Europe

M. I. Al-Jumaili, V. Rauhala, K. Jonsson, R. Karim and A. Parida

Abstract Increased environmental awareness in industry combined with the globalized market economy has created an increase in demand for sustainable and efficient resource utilization. In this context, maintenance plays a critical role by linking business objectives to the strategic and operational activities aimed at retaining system availability performance, cost-efficiency, and sustainability. Performing maintenance effectively and efficiently requires corresponding infrastructure for decision-support provided through eMaintenance solutions. A proper eMaintenance solution needs to provide services for data acquisition, data processing, data aggregation, data analysis, data visualization, context-sensing, etc. For Quality of Service (QoS) in eMaintenance solutions, the performance of both system-of-interest, enabling systems, and related processes have to be measured and managed. However, the QoS has to be considered on all aggregation levels and must encompass the aspects of Content Quality (CQ), Data Quality (DQ), and Information Quality (IQ). Hence, the purpose of this paper is to study and describe some aspects of DQ in eMaintenance related to the process industry in northern Europe.

M. I. Al-Jumaili (✉) · R. Karim · A. Parida
Division of Operation and Maintenance Engineering,
Luleå University of Technology, 971 87 Luleå, Sweden
e-mail: mustafa.aljumaili@ltu.se

R. Karim
e-mail: ramin.karim@ltu.se

A. Parida
e-mail: aditya.parida@ltu.se

V. Rauhala
Kemi-Tornio University of Applied Science, Technology, 941 01 Kemi, Finland
e-mail: ville.rauhala@tokem.fi

K. Jonsson
Department of Informatics, Umeå University, 901 87 Umeå, Sweden
e-mail: katrin.jonsson@informatik.umu.se

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_5,

5.1 Introduction

Since early 2000, development of ICT has contributed to the emergence of the eMaintenance concept which has become a common term in the maintenance related literature today [1]. eMaintenance can be defined as a maintenance strategy where tasks are managed electronically using real-time equipment data obtained through digital technologies (i.e., mobile devices, remote sensing, condition monitoring, knowledge engineering, tele-communications, and internet technologies [2]. eMaintenance is also considered as a maintenance plan to meet the productivity through condition monitoring, proactive maintenance, and remote maintenance through real time information for decision making. eMaintenance thus can be termed as a maintenance support for the e-operation through remote diagnostics and asset management, simulation for optimization and decision making under an e-business scenario for an organization. Under a provider and consumer relationship scenario, the term eMaintenance is often related to providing maintenance services remotely, with the aid of Information and Communication Technology (ICT), which is almost equal to remote support services, [3], through the information logistics.

Information logistics has previously been introduced in the eMaintenance literature to put attention to the flow of information in maintenance processes. Data quality in maintenance forms the integral part of information logistics which is the backbone of a maintenance performance measurement system derived from a clear maintenance strategy that in turn should be derived from and linked to the corporate strategy. The corporate objectives should be cascaded into team and individual goals through a hierarchical system [4], and information logistics. However, current research is limited to exploring data quality issues in specific phases of the maintenance process. This paper takes a broader perspective and contributes by exploring data quality issues in all phases of the maintenance process. The underlying research was conducted as qualitative case studies at a number of Scandinavian processing industries, where the relevance for data quality issues in maintenance was explored. It was found that data quality issues affected all phases of the maintenance process and that a wide variety of quality issues are required to be stressed. Since, prior work has mainly focused on issues with impact on parts of the maintenance process, further investigation of data quality in maintenance are needed.

5.2 Theoretical Framework

eMaintenance is a relatively new and emerging area of research for academia and industry. Initially, the research work was started as a literature study, after which the related issues are identified and studied in the industrial settings. The research work is supported by one organization; in the mining and minor studies with other mining; and paper and pulp industry. The information logistic issues are studied, data collected, and analyzed.

5.2.1 eMaintenance

Maintenance is the combination of all technical and administrative actions, including supervision, intended to retain an item in, or restore it to, a state in which it can perform a required function [5]. A generic maintenance process consists of phases for management, support planning, preparation, execution, assessment, and improvement [5].

The management of maintenance consists of activities as: developing and updating the maintenance policy, providing finances for maintenance and coordinating, and supervision of maintenance. The elements of maintenance support planning are: maintenance support definition, maintenance task identification, maintenance task analysis, and maintenance support resources. Maintenance preparation concerns the planning for specific maintenance tasks, which includes planning of maintenance tasks, scheduling activities, and assigning and obtaining resources. The maintenance execution phase includes the actual performance of maintenance, recording results and special safety and environmental procedures. Maintenance includes measurement of maintenance performance, analysis of results, and assessment of actions to be taken. Finally, maintenance improvement is achieved by improving the maintenance concept, improving resources, improving procedures, and modifying equipment.

The development of IT contributed in the early 2000s to the emergence of eMaintenance, which today is a common term in maintenance-related literature [6]. However, there are different views of what eMaintenance actually is. Some consider eMaintenance as a maintenance strategy, where tasks are managed electronically by the use of real-time item data obtained through digital technologies [2]. Others view eMaintenance as a support to execute a proactive maintenance decision process [6]. In this paper, we adopt a broad view on eMaintenance in line with [6], as maintenance processes managed and performed via computing.

eMaintenance has gained increased attention from research with emphasis on a variety of issues. Muller et al. [7] identifies four general issues addressed in eMaintenance literature: (1) standards, (2) platform development, (3) process formalization, and (4) system development and implementation. Besides these issues the literature also addresses eMaintenance from an information logistics perspective [2, 8, 9]. Maintenance processes are supported by heterogeneous resources such as documentation, personnel, support equipment, materials, spare parts, facilities, information, and information systems [10]. Hence, the provision of the right information to the right user with the right quality and in the right amount of time is essential [4, 11, 12]. This desirable situation can be achieved through appropriate information logistics, which aims to provide just-in-time information to targeted users and to optimize the information supply process.

5.2.2 Data Quality

Quality can be defined as fitness of an artifact's (e.g., product or service) performance to its specification and intended use, from a customer perception [13]. Hence, DQ in eMaintenance need to be considered from a consumer perspective and in the context of maintenance.

The DQ handled by information logistics is an important aspect to consider, since without control of the data quality there will be no control of the output [14]. There are a number of theoretical frameworks for understanding data quality [15–17]. However, to be able to manage data quality you first have to understand what data quality is. A common ground for many of the frameworks is that they involve investigating and describing various categories of desirable attributes of data. The data quality metrics can then for example be used in developing data quality audit guidelines or for use in specification of information systems [17]. Different techniques to improve the data quality have also been suggested, such as data profiling, data standardization, linking, and data cleaning. Data quality is thus of importance both when designing a maintenance information logistic solution and during its operational phase.

However, the quality of the data in manufacturing has not been thoroughly discussed in prior research [18]. Moreover, only a limited number of papers focus on data quality and areas of maintenance, for example maintenance of roads [19] and the data transfer between maintenance workers and facility management [20]. These papers mainly focus on data quality issues with impact on the maintenance execution phase.

In this paper we will adopt the data quality dimensions suggested by Wand and Wang [17] to classify the data quality problems found in our case study: Wand and Wang [17] identifies four types of data quality categories; intrinsic, contextual, representational, and accessibility.

- *Intrinsic* data quality denotes that data have quality in their own right. For example, mismatches among sources of the same data are a common cause of intrinsic data quality problems.
- *Contextual* data quality highlights the requirement that data quality must be considered within the context of the task at hand. This could refer to issues concerning missing (incomplete) data, inadequately defined or measured data, and data that could not be appropriately aggregated.
- *Representational* data quality emphasizes that high quality data should be clearly represented. Examples of issues related to this are interpretability, ease of understanding, concise representation, and consistent representation.
- *Accessibility* data quality refers to the fact that the data must be accessible to the data consumer in a secure way. Accessibility data quality problems are for example characterized by underlying concerns about technical accessibility, data-representation issues interpreted by data consumers as accessibility problems, and data-volume issues interpreted as accessibility problems.

5.3 Method

This paper employed a qualitative case study to explore data quality issues in all phases of the maintenance process. The main corpus of the empirical data for this paper was collected in a case study at a Scandinavian mining company. Besides the main case study, a number of additional case studies were conducted at an additional mining company and five pulp and paper companies in northern Europe. These studies were used as reference studies to see if the major case was unique or not. The findings presented in the paper were identified in all studies.

As a first step, documentation provided by the mining company was studied to understand their organizational structure and the operational process, with a goal to learn about the maintenance process. This study was followed by two rounds of interviews with focus on information/data flows within maintenance, the systems supporting these flows and problems related to these areas. In the first round, six interviews were conducted with maintenance and production personnel. The aim was to select respondents that covered all parts of the maintenance process, as well as different levels of organization and different tasks within that level of organization. In the second round, four interviews were conducted with technicians to understand the information flows and IT support for maintenance execution.

The interviews were semi-structured with open-ended questions and lasted for about 1 h. The interviews structure was based on the subphases of the maintenance process, see Fig. 5.1. These themes guided the semi-structured interviews, but were open to pursuing follow-up questions and issues raised by the respondents. All interviews were tape-recorded and followed up during the data analysis.

After the interviews, the problems related to information flows in maintenance were compiled and structured according to their relevance to data quality and impact on the different phases in the maintenance process, see Table 5.1.

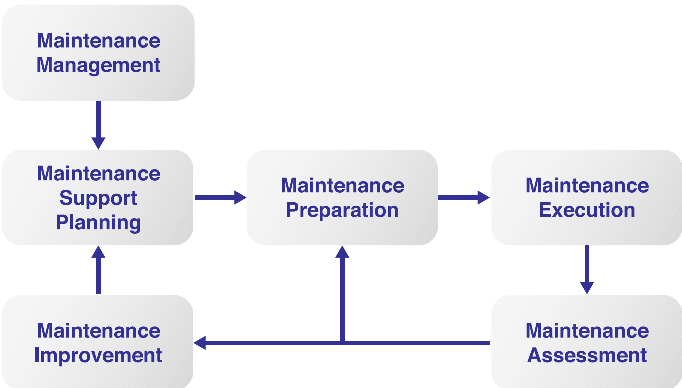


Fig. 5.1 Phases of an overall maintenance process, adopted from International Electrotechnical Commission [5]

Table 5.1 Identified data quality issues and their linkage to maintenance process

Data quality problem	Maintenance management phase	Maintenance support planning phase	Maintenance preparation phase	Maintenance execution phase	Maintenance assessment phase	Maintenance improvement phase	Type of data quality problem
Multiple information sources in spare parts			√				Intrinsic
Lacking progress information for spare part orders		√	√				Contextual accessibility
Manually inputted work orders varies in quality			√	√	√		Intrinsic contextual
Information needed is not connected to work orders				√			Contextual accessibility
Lack of information needed to follow-up work					√	√	Intrinsic contextual
Multiple failure notices			√	√	√		Intrinsic
Usability of systems	√	√	√	√	√	√	Representational
Information duplication and poor maintenance of information		√	√	√	√	√	Intrinsic
Performance indicators missing or difficult to obtain	√				√	√	Contextual
Bad information logistics between teams			√	√			Accessibility
Poor support for work planning by system		√	√				Representational accessibility
Route cause analysis poorly utilized					√	√	Accessibility

5.4 Data Quality Issues Within eMaintenance

The major case study was conducted at a Scandinavian mining company with approximately 160 employees. The plant consists of the actual mine and a concentrator plant. This case study focuses on the maintenance process for the concentrating plant, since maintenance activities in the mine are mainly conducted by subcontractors. The concentrator process at the mine is led by a manager, who has two section managers, one responsible for production and one for plant maintenance and automation. The maintenance and automation function is divided in a mechanical and an electrical team. In our study, we sought to interview people from different positions to get as many perspectives on the maintenance process as possible.

The software (e.g., Computerized Maintenance Management System—CMMS) used in the maintenance process are with few exceptions a number of “stand alone” systems, with poor integration and data exchange. Most of the data have to be manually transferred between the different systems. In addition to the mentioned systems, a large proportion of the maintenance information flows are supported by phone. During the interviews, the aim was to get an understanding of how the different maintenance phases were supported by IT and the information logistics were carried throughout the phases. In the study, a number of problems were identified that is rooted in information logistics and data quality.

5.4.1 Identified Problems

In the case study we have identified a number of problems related to different aspects of data quality. These problems are presented one by one as follows.

5.4.1.1 Data multiplicity

This problem concerns the duplication of information and its existence in different systems or parts of a system. Certain information regarding technical documentation and drawings are stored both in the maintenance system and in other media (i.e., computer hard-drives, CD-ROMs, paper). However, in many cases there is no technical information, drawings, or work instructions in the maintenance system. The stored information is not regularly updated or maintained, thus making it difficult to know if it can be trusted. Sometimes, the information sources contain conflicting information, which makes it difficult to trust. Earlier, the company had a master-student relationship, when new technicians were inducted and information was shared. With the use of a maintenance system, the organization relies upon the documented information and the master–student information sharing relationship was removed. However, as the information in the systems is not valid, the information sharing is not well functioning. The issue with data duplication is worsened by the fact that the responsibility for updating the different data sources is not clear and therefore, seldom done.

Moreover, with the introduction of the maintenance system information related to certain equipment is stored in different subparts of the system. An historical overview of previous maintenance activities is thus difficult to receive. Before the use of a digital maintenance system, the maintenance team had paper device cards that contained technical information, drawings, work instructions, and historical information about previous maintenance actions on certain equipment. One person was responsible for updating these cards which made the information reliable. This information was valuable in trouble shooting and maintenance execution.

5.4.1.2 Manual Input and Transfer of Data

This problem relates to the practice of feeding the systems with information. The practice of manually entering data has led to the data being of varying quality. As an example some of the maintenance personnel enter very detailed and descriptive failure reports in the maintenance system, while others even omit compulsory entries. The maintenance foreman thus often has to consult the operational staff and do a physical inspection to get additional information. Sometimes, the failure reports are very descriptive with descriptions of the failure and potential sources. However, often they just contain a short description of the failure.

Moreover, each time data is manually transferred between systems there is also the risk of data being lost or corrupted. This makes it hard to trust the data and valuable time is therefore lost on finding additional information.

5.4.1.3 Usability of Systems

This problem points to the interface and functionality of the systems used in the maintenance processes. One opinion that many of the interviewees shared was that the usability of the maintenance systems could be improved. It is difficult to find relevant information and guidance for the user on how to use the systems, which are lacking as well.

Furthermore, the maintenance system does not contain any guidance or help for the user, so it can be difficult for the user to know how certain functions should be used or what information should be provided in different fields. The search function in the system is not well developed and makes it difficult to find relevant information, for instance historical information on maintenance actions.

5.4.1.4 Performance Indicators Missing or Difficult to Obtain

To continuously follow-up and improve the maintenance organization, it is important to have indicators of its status and efficiency. Maintenance assessment is currently based on information from SAP, where material costs, salaries for the maintenance staff, and costs for subcontractors are followed up. However, material costs can only be followed on a sub process level, not on equipment level. Moreover, salaries are only followed for the whole plant. There is no information available on the efficiency of the maintenance organization (due to invalid data of work time execution and status of work orders). Thus, it becomes difficult to make assessments and improvements of the organization based on real performance.

5.4.1.5 Poor Information Logistics Between Teams

This problem relates to the transfer of information between different maintenance teams. When, an operator makes a failure notice, he directs it either as a mechanical fault or as an electrical. Some failures are however, not rooted in either a mechanical or an electrical problem, they can only be solved, when electrical and mechanical technicians get together. However, what happens is that the work order is perceived by one of the groups and as they cannot solve the problem, they direct it to the other group, if they cannot find it they direct it back, and so on. As the two groups have separate coffee rooms and different foremen; the information sharing between the groups are occasional and not facilitated by the system or the organizational structure.

5.4.1.6 Knowledge Recycling

This problem concerns the issue of not making use of previous knowledge in the maintenance work. All unplanned shutdowns lasting for more than 1 h are followed up in a root cause analysis. A work template exists for these analyzes with a number of questions that should be answered. The analysis is saved in a word file, which is sent to the section manager who reads and stores it. As the form uses a word file and the storage in a map, these analyzes are difficult to access and the old files are rarely consulted or followed up to find systematic or recurring problems.

Another example of when knowledge is not reused is related to the work orders. When a work order has been completed, it is checked out and a short description of the action is written (example, pump has been replaced). Additional information can be stored, for instance work time and what has caused the failure, but this information is usually not added to the completed work order. Thus, it is difficult to use the work orders as a source of knowledge next time, when an error occurs in the same equipment.

5.4.2 Data Quality in the Maintenance Phases

The perceived problems which were presented in the previous section are all rooted in different aspects of data quality. Table 5.1 gives an overview of the different problems identified during the studies and also linkage between the problems and the phases in the maintenance process. In addition, Table 5.1 also provides a classification of the problems based on the four categories of data quality presented by Wand and Wang [17].

The problem of data multiplicity is classified as an intrinsic problem, since the accuracy of the data may be affected due to the problem of proving the correctness of the data. The issue of manual input and transfer of data is an intrinsic problem, since the accuracy also may be affected by the manual processes. But, it can also constitute a contextual problem since the completeness of the data may be affected as well. The problem related to the usability of systems is a representational problem, since the data may be accurate; but it is difficult to utilize due to poor usability of the system. The problem of performance indicators missing is classified as a contextual problem due to the fact that a lot of data is missing and is not complete. Finally, poor information logistics between teams and root cause analysis difficult to access both is due to accessibility problems since the data is difficult to access. As we can see in Table 5.1, all maintenance phases have a number of data quality problems that are important to address in order to improve the maintenance process.

5.5 Conclusions

Effective and efficient maintenance requires a proper information logistics, which can be delivered through eMaintenance solutions. Development of eMaintenance solutions faces extensive challenges. Some of these challenges can be related to services aimed for acquisition, processing, analysis, and delivery of content aimed for a target context. Others can be related to the content, which is provided though the services. In the both cases, ensuring the quality is highly important. Hence, eMaintenance solutions should provide mechanisms that measure, manage, and visualize QoS and DQ. The conducted case studies indicate that DQ can be related to different features (e.g., usability, accuracy, and relevancy). These DQ features need to be visualized in order to be measured and managed by the eMaintenance solution and visualized by the user.

It can be concluded that, there exist some generic contributions which deals with DQ, but there is a need of developing quality assurance mechanisms within the eMaintenance. These mechanisms should encompass all phases of the maintenance process; be able to manage DQ from different data sources and at various aggregation levels; to visualize DQ for the user; be able to adapt to user's context and the decision process; besides, criticality management. Furthermore, it is also important that there is an overarching plan for information management that covers a system's whole lifecycle.

References

1. Muller A, Suhner MC, Iung B (2007) Maintenance alternative integration to prognosis process engineering. J Qual Maint Eng 13(2):198–211
2. Tsang AHC (2002) Strategic dimensions of maintenance management. J Qual Maint Eng 8(1):7–39
3. Karim R (2008) A service-oriented approach to eMaintenance of complex technical systems. PhD Thesis. Luleå, Luleå University of Technology, Sweden
4. Parida A, Kumar U (2006) Maintenance performance measurement (MPM): issues and challenges. J Qual Maint Eng 12(3):239–251
5. International Electrotechnical Commission (2004) Dependability management—part 3–14: Application guide—maintenance and maintenance support. International Electrotechnical Commission, Geneva, Switzerland, 24 Mar 2004. Report No.: 60300-3-14
6. Kajko-Mattsson M, Karim R, Mirijamdotter A (2010) Fundamentals of the eMaintenance concept. In: Proceedings of the 1st international workshop and congress on eMaintenance, Luleå University of Technology, Luleå, Sweden, 22–24 Jun 2010, pp 147–154
7. Muller A, Crespo Marquez A, Iung B (2008) On the concept of eMaintenance: review and current research. Reliability Eng Syst Saf 93(8):1165–1187
8. Koc M, Lee J (2001) A system framework for next generation EMaintenance Systems. In: Proceedings of the 2nd international Symposium on environmentally conscious design and inverse manufacturing, Tokyo, Japan, 11–15 Dec 2001
9. Parida A, Kumar U (2004) Managing information is key to maintenance effectiveness. In: Proceedings of the intelligent maintenance systems (IMS) international conference. Arles, France, 15–17 July 2004
10. International Electrotechnical Commission (2008) System engineering—system life cycle processes, International Electrotechnical Commission, Geneva, Switzerland. Report No.: 15288
11. Lee J, Ni J, Djurdjanovic D, Qiu H, Liao H (2006) Intelligent prognostics tools and eMaintenance. J Comp Ind 57(6):476–489
12. Parida A, Phanse K, Kumar U (2004) An integrated approach to design and development of e-maintenance system. In: Proceedings of VETOMAC-3 and ACSIM-2004, Allied Publishers Ltd, India, 6–9 Dec 2004, pp 1141–1147
13. Juran JM, Blanton GA (1999) Juran's quality handbook. McGraw-Hill, New York
14. Karim R, Kajko-Mattsson M, Söderholm P (2008) Exploiting SOA within eMaintenance. In: Proceedings of the 30th international conference on software engineering (ICSE), the 2nd international workshop on systems development in SOA environment, Liepzig, Germany, 10 May 2008
15. International Organization for standardization (2009) Data quality—part 100: master data: Exchange of characteristic data: Overview. Report No.: 8000-100:2009
16. Price R, Shanks G (2005) A semiotic information quality framework: development and comparative analysis. J Inf Technol 20(2):88–102
17. Wand Y, Wang RY (1996) Anchoring data quality dimensions in ontological foundations. Commun ACM 39(11):86–95
18. Harding JA, Shahbaz M, Srinivas S, Kusiak A (2006) Data mining in manufacturing: a review. J Mfg Sci Eng 128(4):969–976
19. Bogus SM, Migliaccio GC, Cordova AA (2010) Assessment of data quality for evaluations of manual pavement distress. Geol Prop Earth Mater 2010:1–8
20. Ergen E, Akinci B, East B, Kirby J (2007) Tracking components and maintenance history within a facility utilizing radio frequency identification technology. J Comput Civ Eng 21(1):11–20

Chapter 6
Availability Simulation Modeling of a Roasting Process System

Y. I. Kuruppu and M. R. Hodkiewicz

Abstract Availability simulation modeling is a tool used by industry to assist maintenance decision-making. However, it can be time-consuming to conduct because of poor data quality and the complexity of real systems. In this paper, an availability simulation modeling approach that is suitable to the needs of industry is developed which balances the need for (a) an accurate reflection of system availability performance, and (b) efficiency in development. The approach was tested using data from a gold processing plant and considered 802 assets. The approach is used to construct two reliability block diagrams (RBD), a high level RBD at equipment level and a lower level at component level. The component level model indicates that the roasting system availability is at 94 %, within 3 % of the actual plant downtime; compared to 83 % for the equipment level model. These values are inclusive of equipment failures, as well as, corrective, and preventive maintenance tasks. The component model predicts that 50 % of system unscheduled downtime is caused by 14 failure modes. Potential production revenue savings of 460 h per year exists if these failure modes are addressed. The results suggest that the component-level approach is recommended although the time to construct the model was 8 weeks compared with 5 weeks for the equipment level model. This paper makes suggestions for how to improve the efficiency of development using production time-delay data. Usually maintenance work orders are used for process plant availability simulation studies. The alternative data set was found to improve modeling efficiency by approximately 80 %.

Y. I. Kuruppu (✉) · M. R. Hodkiewicz
The University of Western Australia, Stirling Highway, Crawley 6009, Australia
e-mail: kurupy01@student.uwa.edu.au

M. R. Hodkiewicz
e-mail: melinda.hodkiewicz@uwa.edu.au

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_6,

6.1 Introduction

While availability simulation has been promoted by academic literature and software providers as a viable asset management decision support tool, a gap remains between idealized analysis and real industry scenarios. In particular, availability simulation can be time-consuming to conduct because of the poor quality of real data and the complexity of real systems. Consequently, there is pressure on industry modelers to simplify the model construction process in order to improve efficiency. The minimum level of the reliability block diagram is user-specified, and generally depends on the level of accuracy required for the output results [1]. How does this selection impact on model accuracy and the ability to answer questions for which the model is constructed?

Most studies of process plant availability make use of maintenance work order (WO) data to provide equipment failure information. However, this data source is not intended for reliability analysis, and often requires extensive processing time to make it suitable for analysis and modeling purposes. Previous work demonstrates that time-delay accounting data can instead be used for availability simulation modeling with improved analysis time and little impact to results accuracy. The objectives of this project are to (1) determine the minimum acceptable reliability block diagram level to identify critical assets which constrain system availability, (2) evaluate the trade-off between availability simulation modeling efficiency and accuracy, and (3) present an approach for model construction that is suitable to the needs of industry.

6.2 Literature Review

Availability simulation modeling combines the use of RBD system representation with Monte Carlo simulation method to model complex system behavior. The simulation process repeatedly evaluates the reliability block model (RBM) and an estimate of system availability is made after statistical aggregation of the states of multiple trials. This sampling technique allows for various system states and nonlinear component characteristics to be modeled. While RBM, fault tree analysis (FTA) and Markov analysis are deterministic approaches [2–4], availability simulation is a statistical approach. Statistical models typically involve more data handling than deterministic ones [1], and a drawback of availability simulation is its dependency on computing power. Despite this, availability simulation is the most versatile of all assessment methods for handling complex system configurations and time-dependent component behaviors [2].

Some of the earliest applications of Monte Carlo simulation methods to solve reliability problems are in [5–7]. Since then, simulation studies have grown along with advances in computing power. Specialist software has been developed for many computer simulation applications. Consequently, simulation studies are faster to execute and more accessible to general users; this has encouraged industry to incorporate simulation studies in their daily operations, rather than special occasions only [1].

Research and case studies have established the application of availability simulation modeling to maintenance improvement. Borgonovo et al. [8] and Duffuaa et al. [9] demonstrate that simulation modeling advances analytical methods by enabling one to describe multiple aspects of maintenance management. For example, simulation methods can model the impact of maintenance scheduling, labor organization, equipment aging, and obsolescence, which are not easily captured using analytical methods. Barabady and Kumar [10], Gupta and Tewari [11] and Marquez et al. [12] demonstrated the application of availability simulation software to develop maintenance strategies for processing plant systems. Wysocki and Debouk [13] used Monte Carlo simulation methods to investigate safety–critical system configurations, while Mattila et al. [14], Rodrigues et al. [15] and Kang et al. [16] developed availability simulation models to improve maintenance decision-making for military aircraft systems.

However, some sources acknowledge that challenges exist in the translation of academic analysis to real industry applications [17]. In particular, availability simulation modeling can be time-consuming to conduct because of the quality of real data [18, 19] and validation of model accuracy is lacking in industry applications [20]. Heavey and Ryan [21] explain that understanding of the essence of the system being modeled can be lost in the detailed model construction. A business need exists to develop more efficient modeling approaches, while maintaining results accuracy.

Solutions to improve data quality for availability modeling have been investigated. A study by Veron and Hodkiewicz [22] argues that production time-delay accounting data is of better suitability for the purpose of availability modeling than WO data, as a result of direct measurement of system downtime. This paper also demonstrated that availability modeling using time-delay data can reduce model construction time significantly compared to using maintenance WO data.

Simulation modeling can rapidly become time-consuming and costly to conduct with increasing system complexity [1]. Banks et al. [23] suggest that starting with a simple high level model and increasing complexity only when necessary can improve project efficiency. However, due to a lack of model validation against the real system, industry users may unknowingly produce an unacceptable detriment to results accuracy. Therefore, an investigative aim of this paper is to explore the relationship between system complexity and the trade-off between modeling efficiency and accuracy.

6.3 Method

The basic steps of a simulation study are outlined in Banks [1], these are shown in Fig. 6.1.

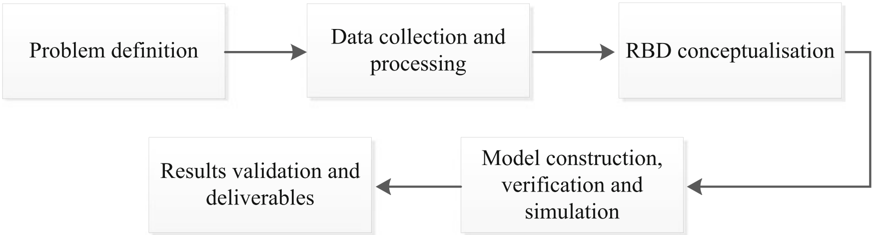


Fig. 6.1 Model construction framework

6.3.1 Problem Definition

According to Jeong et al. [17], “a simulation model is an abstracted representation of a real system to solve specific problems.” The first step, therefore, is to identify the problem and to make an assessment as to whether simulation can provide the desired outputs. The system boundary and project objectives should also be clearly defined. For example, an availability study may aim to answer the following questions:

1. What is the overall system availability?
2. What are the critical assets that constrain system availability?

The solutions to these questions may require different modeling approaches, such as the type of data used; data processing methods; and the granularity of the RBD conceptualization.

6.3.2 Data Collection and Processing

Availability simulation modeling often makes use of maintenance work order (WO) data to provide asset failure information. WOs are a record of maintenance work created by maintenance personnel. They include information about aspects of the maintenance job, such as equipment number, task description, duration; and resources (labor, tools, spare parts). Industry typically has extensive records of WO data, via computer maintenance management system (CMMS). Recently, time-delay accounting data has been more available. This data is characterized by the precise measurement of the duration and location of downtime events. In modern processing plant environments, this production downtime information is generated by programmable logic controllers (PLCs) which automatically log the operational status of a particular asset. The duration between ‘time off’ and ‘time on’ of an asset is called a downtime event, and plant operators are responsible for investigating and providing description of the event in the production downtime database; this description can be related to failure modes.

Next, a decision criterion is followed to prioritise significant asset failure modes within the data. This process intends to reduce the quantity of data used for modeling in order to improve the overall model construction time. The Data Decision Criterion (DDC) proposed by Veron and Hodkiewicz [22] was adapted to this study as follows:

- Repeated events (two or more events with the same equipment and failure mode that occur within <1 h) are merged into a single event. These events are most likely the result of a single failure.
- Similar failures are classified under one broader failure mode. This avoids going into unnecessary detail, for example, 'motor burnt out' and 'motor seized' have the same maintenance action (motor replacement), and hence merging both failure modes as 'motor failure' does not lose relevant information.
- Insignificant failure modes (total sum of downtime <10 h per year) are removed. These events have insignificant effect on overall system availability.
- Events not affecting system reliability are removed (operational activities, external factors outside the system boundary). Non-reliability events cannot be addressed by maintenance actions. The ultimate purpose of the availability study is to aid maintenance decision-making.

Having finalized a set of useable data, the failure data is analyzed to determine underlying distributions. Reliability parameters, time-to-repair (TTR) and time-between-failures (TBF) are calculated for each failure mode, and software tools are used to fit the versatile Weibull distribution to these parameter values. Weibull analysis for this study was conducted using Isograph Limited's Availability Workbench software. Validation of the software-generated distributions is performed by visual inspection of the Weibull plots, goodness-of-fit tests and examination of the correlation coefficient.

6.3.3 RBD Conceptualization

The availability model is graphically represented by a RBD. A good understanding of asset interdependencies, operational knowledge, and maintenance strategy is required in order to create the logical connections of the RBD. Site work is essential in order to validate the RBD with site personnel and obtain anecdotal information [24, 25]. For example, many processing systems will have 'operator work-arounds', whereby downtime is avoided by certain operational actions, or failure of one asset may allow for opportunistic maintenance of other assets. These events are generally not documented, but impact system performance and therefore must be reflected in the availability model.

6.3.4 Model Conceptualization and Verification

Next, the RBD and failure data are integrated into specialized availability simulation software to create the model. This makes simulation studies faster to execute and more accessible to general users. The availability study utilizes Isograph Limited's Availability Workbench AvSim+ module. The user inputs a RBD or fault tree diagram representation of the system. Failure distributions and maintenance tasks are then defined, representing the performance of components within the system. Finally, the model must be verified for consistency in conceptualization.

6.3.5 Simulation

The question of "how long should I simulate, and for how many trials?" Murphy et al. [26] should be addressed for every simulation study. Monte Carlo simulation parameters that must be defined are (a) the forecast period, and (b) the number of iterations. The forecast period defines the domain within which a pseudo-random number generator can select values. These values represent time and are used to sample the probability density functions of the failure distributions. The length of the forecast period should be enough such that it is likely for all outcomes to be sampled at least once [27]. Generally speaking, this means choosing a forecast period greater than the highest mean time between failures (MTBF) of all asset failure modes. There should be enough samples (iterations) to have confidence in the accuracy of the results.

6.3.6 Results Validation

Following simulation, the results are validated against real data to check that the model accurately represents the behavior of the real system. The model is continuously refined until a sufficient degree of validation is achieved. Finally, the required outputs can be extracted from the model and prepared as deliverables.

6.4 Case Study

6.4.1 Problem Definition

The case study is a gold mining operation consisting of a mine and processing plant. One of the operational measures of plant performance is the overall equipment effectiveness (OEE). OEE is a measure of production rate, system availability and product quality:

$$\text{OEE} = \%\text{Production rate} \times \%\text{Availability} \times \%\text{Product quality}. \tag{6.1}$$

Assessment of plant performance has indicated that there may be opportunities for improvement in a sub-system of the processing plant, namely the roasting system. Roasting is a necessary gold extraction process for the ore-type produced at the site. The study focuses on availability improvement of the roasting system to aid in maintenance decision-making. The roasting system is a large and complex sub-system of the processing plant. The model system boundary includes 802 classified assets. External inputs that are necessary to the operation of assets within these areas, such as electrical power, water, and air supply are not considered within the model system boundary.

6.4.2 Data Collection and Processing

Time-delay accounting data was sourced from the production downtime database and exported to Microsoft Excel® spread sheet format. Maintenance WO data was sourced from the CMMS. Data collected from both sources occurred within a 2 year period to the middle of 2010. Due to the large quantity of data, preparation of the data was found to be a time-consuming exercise. Approximately half of the overall project time was spent in this stage.

Data was prepared by (a) classifying failure events, and (b) identifying the equipment failure modes. These steps were heavily reliant on the quality of the description given by the operator responsible for logging the event. Preliminary data analysis indicated that the point-estimated historical availability of the roasting system is 91 %. Unscheduled loss failures account for 85 % of maintenance downtime.

Next, the data decision criterion was applied to (a) select significant unscheduled loss failure modes, and (b) reduce data quantity in order to improve model construction time. In total, 621 failure modes were identified after data preparation of WO data. The DDC filtering process reduced this set to 82 failure modes having significant effect on unscheduled downtime. In comparison to maintenance WO data, production downtime data was found to have a considerably smaller data set with 115 identified failure modes, and 16 after filtering. In both cases, the DDC significantly reduced the quantity of data, to focus on “quick wins”.

Despite the smaller data set, the filtered production downtime data resulted in a comparable Pareto chart to that when using maintenance WO data. Specifically, the top-contributing assets indicated by the production downtime data all appear in the WO data. The noticeable difference between the two is that a greater number of assets contribute to the top 80 % unscheduled downtime when using WO data than production downtime data. The assets that appear in the WO data but not production downtime data analyzes physically exist in standby-duty redundancy configurations. Failure of assets in redundant configurations does not contribute to system downtime (unless simultaneous failures occur), and therefore is not evident in production downtime data. Conversely, failure of assets in redundant configurations is recorded in maintenance WO data, but it is not always known if the failure resulted in system downtime.

From an availability modeling perspective, it is system-level behavior that is of interest. It follows from the failure modes analysis that production downtime data is the more suitable data source. Production downtime data has been shown to identify the same critical assets as WO data, but minus assets which may have large component downtimes but little real effect on system downtime (such as redundancy failure). Furthermore, data processing time using production downtime data was found to take approximately 20 % that of WO data processing, due to the smaller data set size.

6.4.3 RBD Conceptualization

An understanding of the system was developed using process flow diagrams, equipment operating manuals, site area asset list, and information obtained from site personnel. Real system factors such as equipment standby/duty configurations, redundancy switching delay time, storage capacity buffers, equipment ramp times and, etcetera, were considered in the construction of the RBD.

The level of granularity of the RBD was a variable in an investigation of modeling efficiency versus accuracy. Accordingly, coarse and fine-level models were developed by defining the minimum RBD level as major equipment and their components, respectively (Fig. 6.2).

The failure modes identified during the data processing stage were used to define the RBD blocks of the fine model. The coarse model was developed by grouping component failure modes under the parent equipment outage. The latter

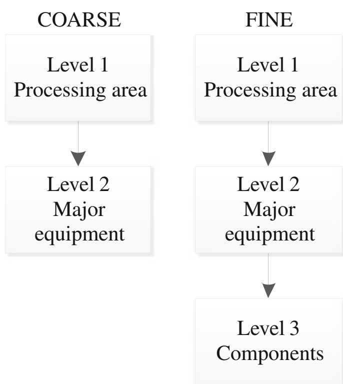


Fig. 6.2 RBD level hierarchy for coarse (*left*) and fine (*right*) models

significantly improves data processing time, as the data isn't required to analyzed to the depth of component failures. Particularly when using maintenance WO data, which can contain numerous events, this approach was found to improve data processing time by approximately 60 %.

6.4.4 Model Conceptualization and Verification

The RBD and failure distributions were integrated in AvSim+ as 'failure models'. Maintenance tasks are assigned to failure models in order to capture the contribution of maintenance time on system availability. During simulation, corrective maintenance (CM) tasks are enacted to return failed assets to operational status, while preventive maintenance (PM) tasks occur at specified fixed-time intervals. Model verification checks connections between RBD blocks; correct assignment of failure models to blocks and the validity of input data.

6.4.5 Simulation

An investigation was conducted in order to determine the optimal number of simulation iterations and forecast period of the availability model. Convergence of predicted availability and downtime occurs at approximately 200 iterations and 6 year forecast (52,560 h). This is greater than the maximum asset MTBF of 3 years.

6.4.6 Results Validation

The output of the simulation stage includes availability predictions and interval spare parts quantities, based on downtime due to unscheduled failures and scheduled maintenance tasks.

Table 6.1 summarizes simulation results.

Table 6.1 Comparison of model RBD granularity results; 6 year simulation period

	Model construction time (wks)	System availability (%)	Scheduled downtime (hrs)	Unscheduled downtime (hrs)
Coarse model	5	83	2,447	11,574
Fine model	8	94	2,672	5,467
System measures	–	91	–	–

6.5 Discussion of Results

6.5.1 RBD Level Granularity

By simplifying the RBD granularity from fine to coarse, an improvement of data processing time of 60 % was achieved. However, results analysis indicates that the coarse model-predicted system availability is underestimated by 8 % (680 h/year) compared with system metric-estimated availability over a 6 year period. The loss of information at the component-level, such as component redundancy and preventive maintenance strategies, results in the model system appearing more unreliable than it is in reality. These factors may account for the underestimation in system availability.

Conversely, the fine model overestimates system availability by 3 % (315 h/year). This error is smaller than that of the coarse model-prediction, and can be attributed to the effect of the DDC, whereby insignificant component failure modes were removed from the data set in order to improve data processing efficiency. The inclusion of these failure modes would see the overestimation in predicted system availability reduced. However, a 3 % error may be justifiable compared with the improved modeling efficiency resulting from the DDC.

Table 6.1 compares downtime as predicted by the coarse and fine models, respectively. It is evident that a significant discrepancy in unscheduled downtime predictions exists; once again, due to the coarse model reflecting a more unreliable system without component-level information. The predicted scheduled downtime (due to PM tasks) is similar for both models, remembering that PM tasks occur at fixed time-intervals in the model and are therefore independent of system reliability.

Similar trends can be observed for the predicted spare parts quantities. The coarse model overestimated the spares quantity by an order of magnitude, while the fine model only slightly underestimated the spares quantity, as compared with actual inventory spares consumption over a 6 year period. Spare parts predictions are particularly sensitive to the lack of identification of component failure modes in the coarse model. During simulation, when a failure event occurs, spare parts are 'used' by the program to return the asset to operational status. When all component failure modes are grouped under the parent equipment outage (as in the case of the coarse model), failure of the equipment will result the program allocating all spare parts for correcting all possible failures, regardless of the actual occurred failure mode. By identifying component failure modes in the data (as in the fine model), spare parts can be selected to address specific failure modes. Therefore, the estimation of spares quantity is significantly improved.

Although reducing RBD granularity to the equipment-level greatly improves the model construction time, it has been demonstrated that it is countered by a significant detriment to results accuracy. Therefore, it is not a suitable approach to modeling processing plant systems for availability simulation, particularly when predicting spare parts quantity.

6.5.2 Model Limitations

In modeling availability of the case study system, equipment failures and maintenance actions were the only contributors to unavailability. In reality, several other factors have significant contribution including spares queuing and labor performance. AvSim+ includes further spares modeling capabilities, including warehouse echelon levels and repair shop features, so the model may be extended to include this as future work. However, the quality of labor performance is more challenging to address, as human reliability modeling poses difficulties [28]. Recently, studies have been made investigating ways of modeling imperfect repairs made by maintenance personnel [29, 30].

Finally, the availability model at present contains no cost evaluation, with the exception of spares purchase costs. Future work will include the costs of labor, tools, spares storage, and transportation.

6.5.3 Time Delay Accounting Versus Maintenance Work Order Data

Approximately half of the overall project time was spent in the data processing stage, due to the large quantity of data for the case study system. Most difficulties arise because data processing, particularly the identification of failure modes, is an iterative process. Notable challenges included:

- Organization of data—significant time was spent on counting/classifying/formatting data, rather than meaningful analysis.
- Classification consistency—different classification of similar data occurred for two major reasons (1) quality of event description and detail given by maintenance personnel (different wording/understanding of the failure event) and, (2) analyst error due to the volume of data.
- Re-processing data—familiarity and better understanding of the data set is developed during data processing. Classifications and assumptions are continuously improved, resulting in data requiring re-processing.
- Outliers—outlying data was judged on their individual merits, rather than being subject to the assumptions of the rest of the data set. This required close attention to all data events in order to identify outliers.

As a comparison to the use of traditional maintenance WO data, production downtime data was analyzed as an alternative data source. Production downtime data was found to be the more suitable data source over maintenance WO data because, (a) it is specifically collected to calculate system-level performance metrics, (b) it identifies the same critical assets as WO data, but minus assets which may have large component downtimes but little real effect on system downtime, and (c) data processing time using production downtime data was found to take approximately 20 % that of WO data processing, due to the smaller data set size.

6.6 Conclusions

Availability simulation studies can be time-consuming to conduct in the typical environment of a site maintenance team, where insufficient time and labor resources are available to conduct a detailed assessment. Understanding of the essence of the system being modeled can be lost in the detailed model construction and simplifications may be made that produce an unacceptable detriment to results accuracy [21]. Literature acknowledges that validation of model accuracy is lacking in industry applications [20]. In this paper, a modeling approach that is suitable to the needs of industry has been developed and validated against data from an industry case study. The case study bridges the gap in the translation of academic analysis to real industry application.

This study suggests that the minimum RBD granularity for process plant systems is component-level, in order to ensure accuracy in simulation results. This study has also confirmed that production time-delay data is of a more suitable quality than maintenance WOs for availability simulation studies, and significantly improves model construction efficiency.

Acknowledgments The authors would like to thank the Company for its support of this project and to the many personnel who gave their time and expertise.

References

1. Banks J (ed) (1998) Handbook of simulation. Wiley, New York
2. Pukite J, Pukite P (1998) Modeling for reliability analysis. IEEE Press, New York
3. Rausand M, Høyland A (2004) System reliability theory: models, statistical methods, and applications, 2nd edn. Wiley, New Jersey
4. StandardsAustralia (2008) Analysis techniques for system reliability—reliability block diagram and Boolean methods. Sydney
5. Gilmore DB (1968) Application of a 'Monte Carlo' method to the determination of the mean time to failure of complex systems. In: IEEE Proceedings of the 1968 annual reliability symposium. pp 391–396
6. Hall HD, Ness WG (1970) Maintenance and reliability simulation modelling for air freighters. Ann Reliab aintainability 9:354–361
7. Absher Jr GW (1971) Reliability approach to ship systems. In: IEEE Proceedings of the 1971 annual reliability symposium, p 194
8. Borgonovo E, Marseguerra M, Zio E (2000) A Monte Carlo methodological approach to plant availability modeling with maintenance, aging and obsolescence. Reliability Eng Sys Saf 67:61–73

9. Duffuaa SO, Ben-Daya M, Al-Sultan KS, Andijani AA (2001) A generic conceptual simulation model for maintenance systems. J Qual Maint Eng 7(3):207–219
10. Barabady J, Kumar U (2007) Reliability analysis of mining equipment: a case study of a crushing plant at Jajarm Bauxite Mine in Iran. Reliability Eng Syst Saf 93:647–653
11. Gupta S, Tewari PC (2009) Simulation modeling and analysis of a complex system of a thermal power plant. J Ind Eng Manag 2(2):387–406
12. Marquez AC, Heguedas AS, Iung B (2005) Monte Carlo-based assessment of system availability: a case study for cogeneration plants. Reliability Eng Syst Saf 88:273–289
13. Wysocki J, Debouk R (2007) Methodology for assessing safety-critical systems. Int J Model Simul 27(2):99–106
14. Mattila V, Virtanen K, Raivio T (2008) Improving maintenance decision making in the Finnish air force through Simulation. Interfaces (Informs) 38(3):187–201
15. Rodrigues MB, Karpowicz M, Kang K (2000) A readiness analysis for the Argentine air force and the Brazilian Navy A-4 fleet via consolidated logistics support. In: WSC 2000 proceedings of the 32nd conference on winter simulation. Orlando, Florida, USA, pp 1068–1074
16. Kang K, Gue KR, Eaton DR (1998) Cycletime reduction for naval aviation deports. In: WSC 1998 proceedings of the 30th conference on Winter Simulation. Washington, DC, USA, pp 907–912
17. Jeong KY, Wu L, Hong JD (2009) IDEF method-based simulation model design and development. J Ind Eng Manag 2(2):337–359
18. Ballou D, Madnick S, Wang R (2004) Assuring information quality. J Manag Inf Syst 20(3):9–11
19. Gouws L, Trevelyan J (2006) Research on influences on maintenance management effectiveness. In: Proceedings of the ICOMS asset management conference, Sydney, Australia, pp 72–79
20. Andijani A, Duffuaa S (2002) Critical evaluation of simulation studies in maintenance systems. Prod Planning Control 13(4):336–341
21. Heavey C, Ryan J (2006) Process modelling support for the conceptual modelling phase of a simulation project. In: WSC 2006 proceedings of the 2006 winter simulation conference. Monterey, CA, USA, pp 801–808
22. Veron B, Hodkiewicz MR (2009) Rapid availability model development using delay accounting data. In: Proceedings of the ICOMS asset management conference. Sydney, Australia
23. Banks J, Carson JS, Nelson BL (1996) Discrete-event system simulation, 2nd edn. Prentice Hall, New Jersey
24. Kortelainen S, Piirainen K, Karkkainen H, Tuominen M (2011) Dependency between learning and profitability in different industry conditions: a system dynamic simulation. Int J Bus Innov Res 5(2):229–253
25. Clemen RT, Winkler RL (1999) Combining probability distributions from experts in risk analysis. Risk Anal 19(2):187–203
26. Murphy KE, Carter CM, Wolfe LH. (2001) How long should I simulate, and for how many trials? A practical guide to reliability simulations. In: Proceedings of the reliability and maintainability symposium. Philadelphia, PA, USA, pp 202–212
27. Sawilowsky S (2003) You think you've got trivials? J Mod Appl Stat Methods 2:218–225
28. Salvendy G (ed) (1987) Handbook of Human Factors. Wiley, New York
29. Doyen L, Gaudoin O (2004) Classes of imperfect repair models based on reduction of failure intensity or virtual age. J Reliab Eng Syst Saf 84(1):45–56
30. Lim JH, Kim DK, Park DH (2005) Cost evaluation for an imperfect-repair model with random repair time. Int J Sys Sci 36(11):717–726

Chapter 7
Maintenance Optimization for Asset Systems with Dependent Performance Degradation

N. Rasmekomen and A. K. Parlikad

Abstract This paper focuses on consider the optimization of maintenance plans for a system of assets with failure/degradation interaction between the assets. The asset system under consideration in this paper has *M* identical non-critical machines feeding their output to a critical machine. A common repair team performs maintenance on all the machines. All the machines deteriorate over time independently. In addition to the independent degradation, the performance of the non-critical machine also affects the degradation of the critical machine. We develop a mathematical model to represent the interactions and performance of the asset system. We also provide a simulation-based numerical solution to optimize the maintenance plan for the system outlining the maintenance intervals for each of the machines.

7.1 Introduction

Complex engineering systems that consist of multiple assets (or complex assets with multiple critical sub-components) have received more attention in the past few decades [1]. For a complex multi-asset system (e.g., a manufacturing line with multiple machine tools and production equipment), replacing or repairing each asset requires understanding its effects on other assets.

N. Rasmekomen (✉) · A. K. Parlikad
Institute for Manufacturing, Engineering Department, University of Cambridge, Cambridge CB3 0FS, UK
e-mail: nr324@cam.ac.uk

A. K. Parlikad
e-mail: aknp2@cam.ac.uk

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_7,

The interactions between assets in multi-asset system can be classified into three types [2].

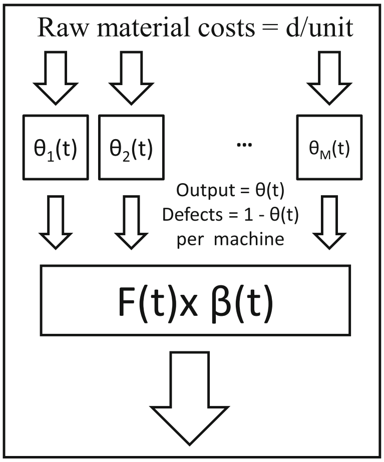
1. *Economic dependence*: This means that the sum of individually maintaining each asset in the system yields different costs (higher or lower) from grouping some of the assets together for joint maintenance. For example, while maintaining traffic control lights, when a maintenance team is sent out to replace a failed light bulb, it is often more economical to take the opportunity to replace other working light bulbs due to the fixed maintenance setup cost includes transportation of manpower and equipment [3].
2. *Structural dependence*: Systems have structural dependence when working components have to be replaced or at least dismantled in order to replace or repair failed ones [2]. These constraints are applied to the maintenance personnel on a decision making of group maintenance activities [4].
3. *Stochastic dependence or failure interaction*: This type of dependence occurs when the condition of one asset in the system influences the state or the failure rate of other assets. For example, in the oil as gas sector when a gas valve is no longer in its operating condition, it could lead to significant pressure-load in a pipeline, which in turn causes the pipe to degrade at a much higher rate. This kind of interaction is common in mechanical systems [5].

The most complete review of the literature in maintenance of complex systems can be found in Nicolai and Dekker [4]. An example of a recent work in this area is from Lai [6] which studied a replacement problem of a two-unit parallel system with Type II failure interaction [4].

Berk and Moinzadeh [7] studied maintenance scheduling of a system with the M machines deteriorating independently. Models were developed for the case of deterministic and random maintenance times. This paper extends this work and studies a complex engineering system that consists of M identical deteriorating non-critical machines and a critical machine. Degradation of the non-critical machines influences degradation of the critical machine. Failure interactions between the non-critical machines and the critical machine are studied in terms of deterministic maintenance times. An industrial example of such a system is a waste-water treatment system where water filters which have waste water as inputs feed their outputs to a critical pump. The performance degradation of the pump would also depend on the quality of outputs from the filters. Lower quality of the outputs from the filters could create sediments on the pump and result in degradation of the pump's performance. The aim is to optimize the maintenance plan for such a system provided that human resource constraints are also in place.

The structure of this paper is organized as follows; the next section outlines the problem definition. Then, Sect. 7.3 provides numerical results of the optimization. Further analysis of the problem to demonstrate the impact of this proposed method is derived in Sect. 7.4 with conclusions in Sect. 7.5.

Fig. 7.1 System structure



7.2 Problem Formulation

Consider a system consisting of M identical parallel machines and a critical machine as demonstrated in Fig. 7.1. We aim to maximize profit per unit time of the system by optimizing time between maintenance for the parallel machines and the critical machines.

The system is fed by raw materials at a rate of one unit per time to each of the parallel machines. Raw material cost is d per unit. All the parallel machines feed the critical machine. Final products are the output of the critical machine.

The assumptions made are the following:

- Ages of the machines refer to 'usage age' of the machine, i.e., the age of a machine does not increase when it is not operating.
- Maintenance actions are assumed to be perfect and the transition time between when a machine is in operation and when it is under maintenance is negligible.
- Maintenance times are constant and deterministic.
- The system is maintained by a single maintenance crew. Therefore, only one machine can be maintained at a time.
- The parallel machines are subject to independent performance degradation only.
- The critical machine is subjected to both dependent and independent performance degradation.
- When the critical machine is under maintenance, the whole system is down.

Variables used throughout the paper are defined as follows:

- M = number of identical parallel machines in the system
- $\theta_i(x)$ = Yield function of an ith ($1 \leq i \leq M$) parallel machine at age x
- $\beta(y)$ = Yield function of the critical machine at age y
- $F(y)$ = Efficiency function of the critical machine, $0 \leq F(y) \leq 1$
- d = Raw material cost per unit

- k = cost of maintaining a parallel machine
- w = cost of maintaining a parallel machine
- $1/\mu$ = amount of time required for maintenance of a parallel machine
- $1/\lambda$ = amount of time required for maintenance of a parallel machine
- τ = amount of time a parallel machine is in operation before taken off for maintenance
- T = amount of time the critical machine is in operation before taken off for maintenance
- n = number of parallel machines maintained before maintenance of the critical machine

The performance of each parallel machine i ($1 \leq i \leq M$) is represented by a yield function, $\theta_i(x)$ where x is an age of the machine. $\theta_i(x)$ is a decreasing function of x with values range from 1 to 0, reflecting the degradation of each machine. Therefore, amount of defects produced by a parallel machine of age x is given by $1 - \theta_i(x)$. Each parallel machine is in operation until its age reaches τ time units before it is taken off the system for maintenance. The time required for each maintenance action is $1/\mu$ time unit and the cost of maintenance is k. While one of the parallel machines is under maintenance, the system keeps operating with the remaining machines. After maintenance, the yield value is restored to 1 and the machine goes back into the system and keeps operating until the next maintenance is required.

The failure interaction of the system is that defective outputs of the parallel machines decrease the performance of the critical machine. The effects of this interaction can be represented by an efficiency function, $F(y)$, where y is an age of the critical machine.

$$F(y) = 1 - b\int_0^y \sum_{i=1}^{M} (1 - \theta_i(x))dx \tag{7.1}$$

where b is a constant positive coefficient of the defect effects ($b \geq 0$). Please note that once $F(y)$ is 0, the critical machine is effectively down.

While the critical machine's efficiency is reduced due to the accumulated defects of the parallel machines' outputs, independent degradation of the critical machine is reflected in its yield function, $\beta(y)$. $\beta(y)$ is a decreasing function of y with values range from 1 to 0.

The final products of the system, $Z(t)$, at time t can then be written as:

$$\begin{aligned} Z(t) &= \beta(t)F(t)\left(\sum_{i=1}^{M} \theta_i(t)\right) \\ &= \beta(t)\left(1 - b\int_0^t \sum_{i=1}^{M} (1 - \theta_i(x))dx\right)\left(\sum_{i=1}^{M} \theta_i(t)\right) \end{aligned} \tag{7.2}$$

In addition to the maintenance on the parallel machines, the critical machine would also need to be maintained with a maintenance time of $1/\lambda$. The cost of maintaining the critical machine is w. After this maintenance, the age of the critical machine is reset to 0 and $\beta(y)$ is restored to 1.

Clearly, a high value of τ would result in a lower maintenance cost but a higher rate of defects going through to the critical machine. Low values of τ would result in a lower number of final products available due to machine maintenances, which also affect the throughput. τ can then be seen as a decision variable (maintenance policy for the parallel machines), which reflects the limits in terms of production requirements and product quality.

The number of the parallel machines in operation at a time is also determined by the value of τ. For the case where $\tau < (M{-}1)/\mu$, there are either j or $j{-}1$ parallel machines in operation at a time, where $j < M$. As a result, there are either $M - j$ or $M - j + 1$ machines either waiting or under maintenance, which means the maintenance resources are always busy. For the case where $\tau > (M{-}1)/\mu$, there are either M or $M{-}1$ parallel machines in operation at a time. Therefore, the maintenance resources are not always busy. For a case where $\tau = (M{-}1)/\mu$, there are always $M{-}1$ machines in operation and the maintenance resources are always busy since there is always a machine under maintenance (none waiting). Figure 7.2 demonstrates the effects of different values of τ when there are four parallel machines ($M = 4$). From Fig. 7.2, under steady-state conditions, it can be observed that the ages of the operating machines are exactly $1/\mu$ apart due to their different starting point of the operation after maintenance. Therefore, the yield function of a parallel machine $r + 1$ ($\theta_{r+1}(t)$, $1 \leq r < M$) is a time-shifted function of r/μ time units from the yield function of a certain parallel machine ($\theta_1(t)$) in the system's time domain.

The maintenance policy for the critical machine is to decide how long the critical machine would have to be in operation (T) before it undergoes maintenance. This can also be measured in terms of the number n of the parallel machines undergoing maintenance before the critical machine.

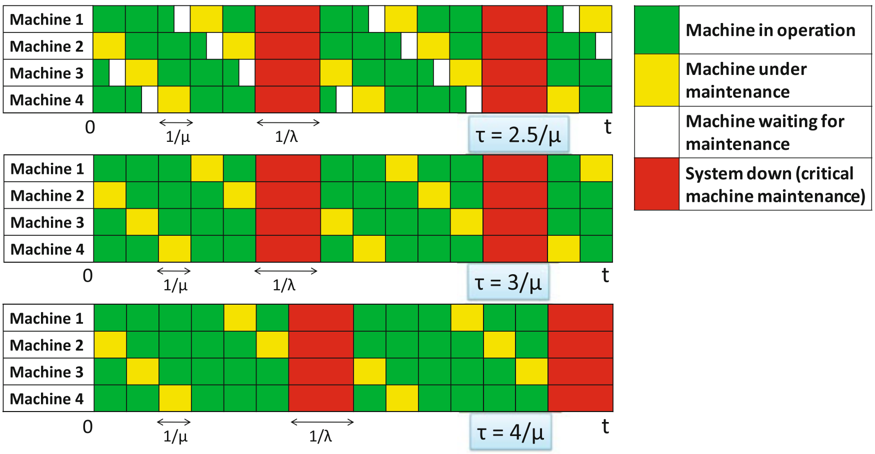


Fig. 7.2 Operating schedule of a system with 4 parallel machines and a critical machine at different values of τ ($n = 5$)

It can be observed from Fig. 7.2 that when $\tau > (M-1)/\mu$, there is a period of $\tau - (M-1)/\mu$ where there is no maintenance action at all. Therefore, the average amount of time until nth maintenance of the parallel machines can then be given by:

$$T(n) = \begin{cases} \frac{n}{\mu} & \text{when } \tau \leq (M - 1)/\mu \\ \frac{n}{\mu}\left(\frac{\mu\tau + 1}{M}\right) & \text{when } \tau > (M - 1)/\mu \end{cases}. \tag{7.3}$$

Once the maintenance policies on both the critical machine and the parallel machines are determined, the system can be viewed as if it has an operating cycle of $T + (1/\lambda)$. The optimization problem is then to find an optimal value of τ and also an optimal number of maintenances (n) of the parallel machine before maintaining the critical machine to maximize the profit of the system.

For a final product price of 1 per unit, the profit per unit time of the system, $P_t(n, \tau)$, can be given by:

$$P_t(n,\tau) = \frac{\int_0^T \beta(t)\left(1 - b\int_0^t \sum_{i=1}^M (1 - \theta_i(x))dx\right)\left(\sum_{i=1}^M \theta_i(t)\right)dt - \{(k \times n) + w\} - (d \times n \times \tau)}{T + \frac{1}{\lambda}}. \tag{7.4}$$

The next section provides a numerical solution for optimizing τ and $T(n)$ which would maximize P_t.

7.3 Numerical Example

This section provides some numerical results of this study. Please also note that the analytical result is also studied in a special case where $\tau = (M-1)/\mu$. This result is not provided in this paper but will soon be published.

Numerical values used to simulate the results in this section are:

$$b = 0.01, \mu = 1, \lambda = 0.5, M = 4, w = 0.5, k = 0.1, \tau = 3, d = 0.2$$

$\theta_i(t)$ and $\beta(t)$ are represented by exponential functions $e^{-0.1(t-\frac{i-1}{\mu})}$ and $e^{-0.01t}$ respectively. Figure 7.3 represents the output yield function of one parallel machine at time. The machine started operation at $t = 0$ with maximum output yield of 1. After τ time units of operation, it took $1/\mu$ time unit to undergo maintenance and then went back to operation at the maximum output yield. Figure 7.4 demonstrated the output yield of each different parallel machine and also the total output of the machines.

Final products are produced in cycles as per maintenance policy of the parallel machines and the critical machine as demonstrated in Fig. 7.5. Since τ is

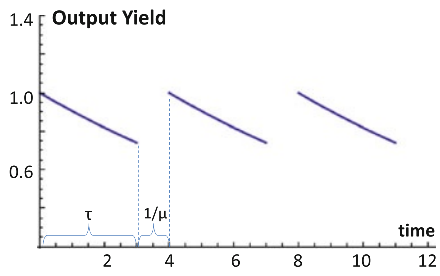


Fig. 7.3 Output yield of a parallel machine, $\theta_I(t)$

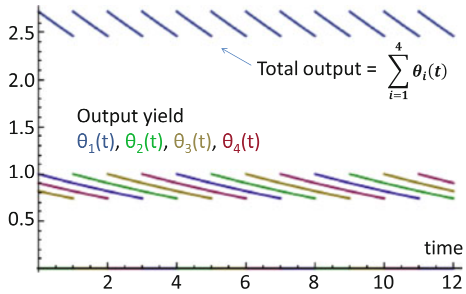


Fig. 7.4 Output yield of the parallel machines and the total output

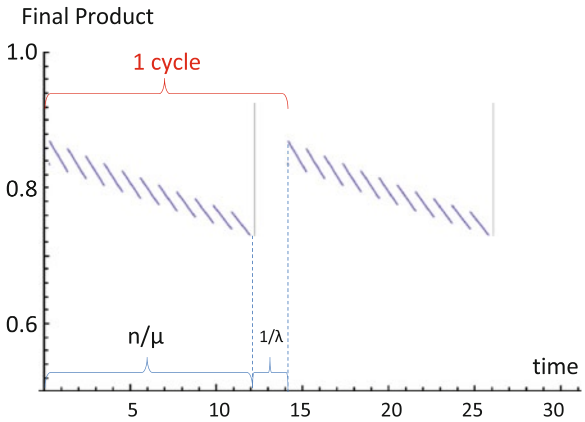
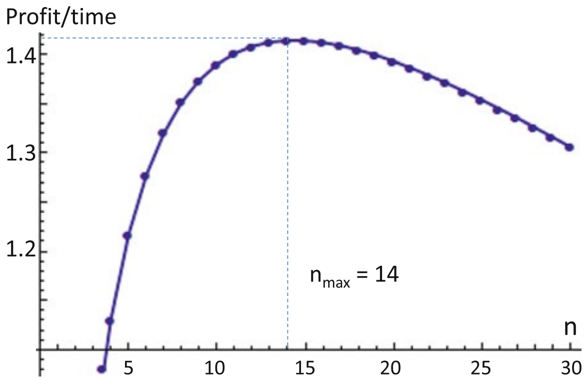


Fig. 7.5 Final products of the system with n maintenances of the parallel machines before a critical machine maintenance

determined as 3 time units, the objective is to find an optimal value of n that will maximize the profit per unit time of the system, $P_t(n, 3)$.

Since $\tau \leq (M-1)/\mu$, therefore $T(n) = n/\mu$. Profit per unit time function as given in Eq. (7.4) can be plotted for different values of n as shown in Fig. 7.6. It can be observed that the optimal n which maximizes the profit per unit in this setting is $n = 14$. On the other hand, if n is predetermined, τ can also be optimized using Eq. (7.4). Since the profit per unit time of the system is a function of τ and n, optimizing the two variables at the same time is also possible.

Fig. 7.6 Profit per unit time of the system at different values of n ($\tau = 3$)



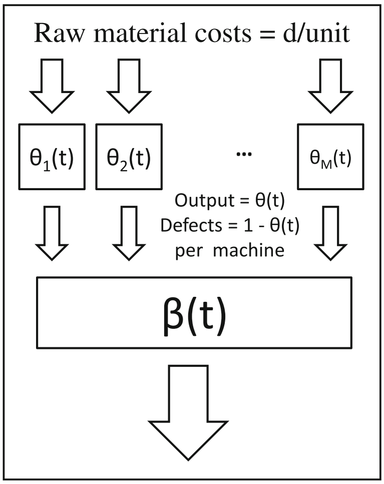
7.4 Discussion

In this section, we discuss why explicit consideration of interactions is important in maintenance planning for complex systems. If we do not take failure interactions, as modeled by $F(t)$, into account, it would result in a system structure as shown in Fig. 7.7. Maintenance planned according to this setting of the system would be different from the original setting in Fig. 7.1. With $F(t)$ missing, profit per unit time function of the system in Fig. 7.7 is given by:

$$P_t^*(n,\tau) = \frac{\int_0^T \beta(t)\left(\sum_{i=1}^{M} \theta_i(t)\right)dt - \{(k \times n) + w\} - (d \times n \times \tau)}{T + \frac{1}{\lambda}} \tag{7.5}$$

With the same set of numerical values used in Sect. 7.3, Fig. 7.8 compares the numerical results of two different models used to find an optimal value of n when $\tau = 3$ (using Eqs. 7.4 and 7.5). Without explicit modeling of failure interactions,

Fig. 7.7 System structure (without failure interactions)



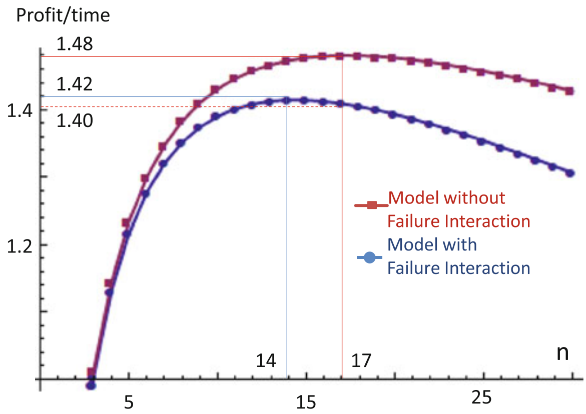


Fig. 7.8 Profit per unit time of the system as modeled in Figs. 7.1 and 7.7 at different values of n ($\tau = 3$)

the maintenance policy for the critical machine would be given by $n = 17$ (instead of 14) with one expecting a profit per unit time of 1.48. However, once the maintenance plan is deployed, the system would only provide the profit per unit time of 1.40. This value is 0.02 lower than the maximum profit/time of 1.42 at $n = 14$. This demonstrates that failure to facilitate the dependencies of the system have resulted in a loss of $\frac{0.02 \times 100}{1.42} = 1.4\,\%$ of the profit/time of the system. Similar kind of analysis can also be applied when optimizing τ for the parallel machines.

Another interesting observation can be made by examining the level of interactions between the parallel machines and the critical machine. According to (1), it can be observed that the higher value of b would result in more impacts of the defective outputs on performance of the critical machine, i.e., more failure interactivities. Figure 7.9 demonstrates Profit per unit time of the system ($P_t(n, 3)$) at different values of b. It can be observed that the optimal values of n which maximize the profit/time of the system vary with values of b. This is due to a fact that the higher values of b mean that the performance of the critical machine is decreasing more rapidly, hence $T(n)$ would also have to be shorter so that the critical machine can be maintained more frequently. As a result, the maximum profit/time of the system would also be less for the high values of b due to more disruptions for the critical machine maintenance.

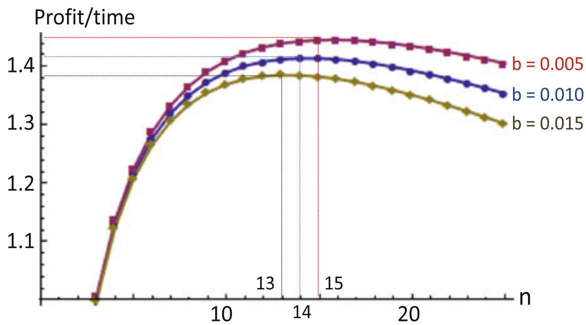


Fig. 7.9 Profit per unit time of the system $P_t(n, 3)$ at different levels of interactivities

7.5 Conclusion

This research aims to contribute to the area of maintenance of complex multi-asset systems. This paper presented a simulation model to optimize maintenance plans for a system that consists of M parallel machines and a critical machine when failure interactions exist. Numerical results demonstrated that the maintenance plan for the M parallel machines and the critical machine can be optimized by the proposed method. It was also showed that failure interactions can have significant influence on maintenance optimization.

References

1. Gurler U, Kaya A (2002) A maintenance policy for a system with multi-state components: an approximate solution. Reliab Eng Syst Saf 76(2):117–127
2. Thomas L (1986) A survey of maintenance and replacement models for maintainability and reliability of multi-item systems. Reliab Eng 16(4):297–309
3. Vab der Duyn Schouten F, Vlijmen BV, Vos de Wael S (1998) Replacement policies for traffic control signals. IMA J Math Appl Bus Ind 9(4):325–346
4. Nicolai RP, Dekker R (2006) Optimal Maintenance of Multi-component Systems: a review. Econometric Institute, Erasmus University, Rotterdam
5. Sun Y, Ma L, Mathew J (2009) Failure analysis of engineering systems with preventive maintenance and failure interactions. Comput Ind Eng 57(2):539–549
6. Lai MT (2009) A discrete replacement model for a two-unit parallel system subject to failure rate interaction. Qual Quant 43(3):471–479
7. Berk E, Moinzadeh K (2000) Analysis of maintenance policies for machines with deteriorating performance. IIE Trans 32(5):433–444

Chapter 8
Prognostics for Optimal Maintenance: Maintenance Cost Versus Product Quality Optimization for Industrial Cases

A. V. Horenbeek, A. Bey-Temsamani, S. Vandenplas, L. Pintelon and B. Deketelaere

Abstract Correlation between the quality degradation of a product and maintenance of a machine is often established by the production engineers. To asses this correlation, some assumptions are made. In most cases it is assumed that the quality of the product degrades after a fixed number of operation cycles of the production machine. Therefore, maintenance of the production machine is only performed after this number of cycles is accomplished. This kind of assumptions is often not valid in modern industry since high variability of products, tolerances of machines/components, reliability variations of these components, extensive/smooth usage, etc., make this degradation quite dynamic in time. As a result, the quality of the product could get degraded in a fast way if this variability is high or in a slow way if this variability is low. Both cases will lead to low benefit because of lost production in the former case or redundant maintenance in the latter one. In this paper, we propose a solution to this problem by maximizing the benefit using online monitoring of product's quality degradation and maintenance cost evolution. A Condition Based Maintenance framework for industry developed in Prognostics for Optimal Maintenance (POM) project [1] and described in [2] is applied to two industrial use cases in order to deploy and validate the proposed technique.

A. V. Horenbeek (✉) · L. Pintelon · B. Deketelaere
Centre for Industrial Management, Celestijnenlaan 300A B-3001 Leuven, Belgium
e-mail: adriaan.vanhorenbeek@cib.kuleuven.be

L. Pintelon
e-mail: liliane.pintelon@cib.kuleuven.be

B. Deketelaere
e-mail: bart.deketelaere@biw.kuleuven.be

A. Bey-Temsamani · S. Vandenplas
Flanders' Mechatronics Technology Centre, Celestijnenlaan 300D B-3001 Leuven, Belgium
e-mail: abdellatif.bey-temsamani@fmtc.be

S. Vandenplas
e-mail: steve.vandenplas@fmtc.be

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_8,

8.1 Introduction

Many models for Condition Based (CBM) or Predictive (PdM) Maintenance optimization exist in literature [3–6]. These models make clear that maintenance decision making based on real-time information from the components and systems has a substantial benefit regarding maintenance cost, prevention of unexpected failures, and reduction of downtime. However, for some systems it is not due to the state of the system itself that maintenance is needed but due to the quality degradation of the products the system is producing. The objective of this paper is to determine the benefit of CBM with regard to quality degradation of the product using a condition monitoring system. By this mean, the optimal time to perform maintenance is determined by considering the trade-off between maintenance cost and cost of quality degradation of the products. Predictive maintenance optimization in literature is mostly restricted to theoretical modeling of the degradation process, for example stochastic processes, and subsequently finding an optimal maintenance policy for this degradation process [7, 8]. This results on putting many assumptions about the failure behavior of components. However, these assumptions are only valid under certain circumstances. Operation and environmental conditions are assumed to be known and cannot change significantly. In general, however, this is not true for all machines because usage rates and environmental conditions are changing over time. These variations are unfortunately often not taken into account for the case studies covered in literature. This creates a gap between theory and practice [3, 9]. Moreover decision making based on real-time information from monitoring systems and components is still an under explored area in maintenance optimization [10]. The integration of predictive information into decision support systems is a very important step that needs further research. To overcome these flaws, two case studies are presented in this paper where real-time predictive information coming from the industrial machines is directly used to support maintenance decision. This support is given by updating a cost function recursively when new information about the system performance becomes available. Based on this information, maintenance is scheduled in an optimal way. The activities of maintenance optimization by using condition-based maintenance are performed within the Prognostics for Optimal Maintenance (POM) project [1]. The POM CBM framework is used as a tool for maintenance optimization by using a CBM policy [2]. This framework is based on the ISO-13374 standard for condition monitoring and diagnostics of machines [11]. The originality and contribution to the field of maintenance of this paper lies at different levels:

A maintenance cost versus product quality degradation CBM optimization is performed, which has, according to the knowledge of the authors, never been published before. Although this should certainly be considered in many industrial production machines in order to be able to perform optimal maintenance. Degradation is not only caused by wear out, but mainly by usage rates and environmental conditions, which is accounted for in the condition monitoring approach taken in this paper.

An integrating approach is taken by developing a decision support system based on condition monitoring information directly coming from the machine without any assumptions on the degradation process. The integration of condition monitoring that results in predictive information on equipment performance and decision making based on this predictive information is perceived as one of the biggest challenges in maintenance [10].

Few case studies have been reported on maintenance optimization models for condition-based maintenance [9]. In this paper, the developed CBM optimization model is applied to two different case studies to show the applicability of the developed methodology on real life industrial cases.

This paper is organized as follows. Section 8.2 describes the two different case studies considered for optimizing the condition-based maintenance policy. The POM CBM framework is described briefly in Sects. 8.3 and 8.4 presents the maintenance optimization model based on the predictive information. Finally, conclusions and future research are stated in Sect. 8.5.

8.2 Use Cases Description

8.2.1 Seal Quality Monitoring in a Packing Machine

The first industrial use case consists on a Vertical Form Fill and Seal (VFFS) packing machine. The machine produces bags of different products (chips, cheese, sugar, etc.) in food industry. A plastic film roll is supplied as a packaging material. After forming flaps that wrap around a main conical tube as depicted in Fig. 8.1, the film is pulled downward around the outside of the tube and a vertical heat-sealing jaws clamp onto the edges of the film bonding the film by melting the seam edges together. After the bonding, a knife cuts the film forming thus a produced bag.

One of the main rejects in the field is the seal quality of the produced bags. The seal quality degradation is caused by accumulation of dirt and dust from the production environment and by leakage of the products during the cutting process on the sealing jaws and reducing their sealing performance. In order to monitor this degradation, a condition monitoring system called SealScope [12] is used. This system measures vibration signatures due to the impacts of sealing jaws during the bonding process and applies advanced multivariate quality control charts technique [13] to calculate prognosis features correlated to the studied degradation.

8.2.2 Print Quality Monitoring in Copiers

The second use case consists on monitoring the quality of the copy papers from a fleet of copiers. This case study was carried out in a previous project called IRIS

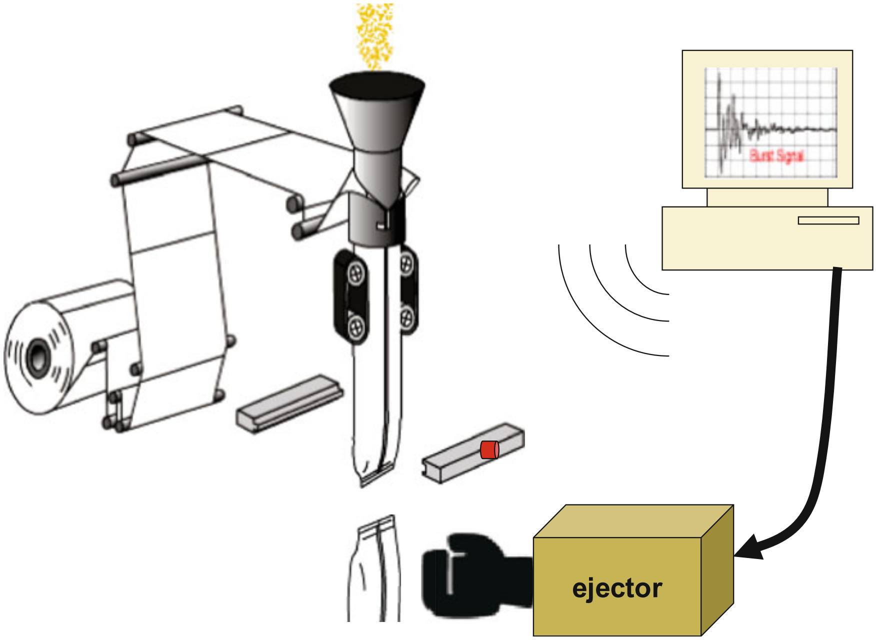


Fig. 8.1 Seal quality monitoring in a packing machine

(Intelligent Remote Industrial Services) [14, 15] where industrial services (machine's health management, remote configuration, etc.) are provided to the customers of machines builders. An illustration is shown in Fig. 8.2. Considering the machine's health management, the main purpose in that project was to identify features which are correlated to the degradation of different components in the copiers and perform predictive maintenance using these predictive features. In this work, the predictive maintenance action will not only be optimized based on degradation of components but also taking into account degradation of produced papers.

8.3 POM CBM Framework for Condition-Based Maintenance in Industry

In order to facilitate the design and the deployment of a Condition-Based Maintenance policy in industry, more precisely machine manufacturer and production machinery, the POM CBM framework has been developed in the frame of POM project [1]. An illustration of the architecture of this framework is depicted in Fig. 8.3.

A detailed description of this framework is given in [2]. The framework has been developed following the available standards like ISO 13374 for condition

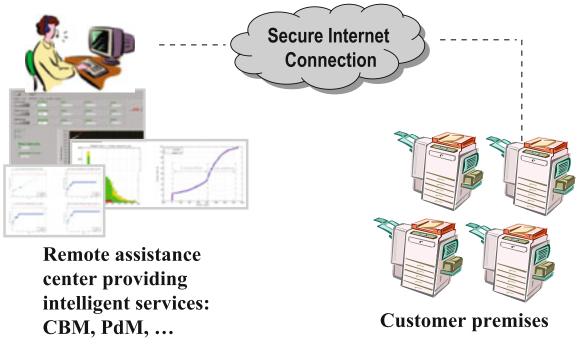


Fig. 8.2 Print quality monitoring in copiers

monitoring and diagnostics and OSA-CBM (Open Systems Architecture for Condition-Based Maintenance) [11].

The different modules of the framework consist on (1) data generation, (2) features generation, (3) modeling and assessment, and (4) advisory generation. Every module is considered as independent from the others and could be customized accordingly to the studied application. This is afforded by a proper choice of inputs/outputs interfacing every module and allowing thus flexibility and interoperability. A module could also be divided to sub-modules where some external interactions could be set, like including information from experts in the field if available to make an assessment model more robust. Since the framework is supposed to work online for maintenance optimization, feedbacks between different modules are foreseen to fortify in an iterative way the prognostics against variability and environmental changes in the studied process.

In next sections, only the last module, Advisory Generation, will be described for the two studied use cases. The details about the predictive features used for the prognostics part are available respectively in [13] and [15] for the two use cases.

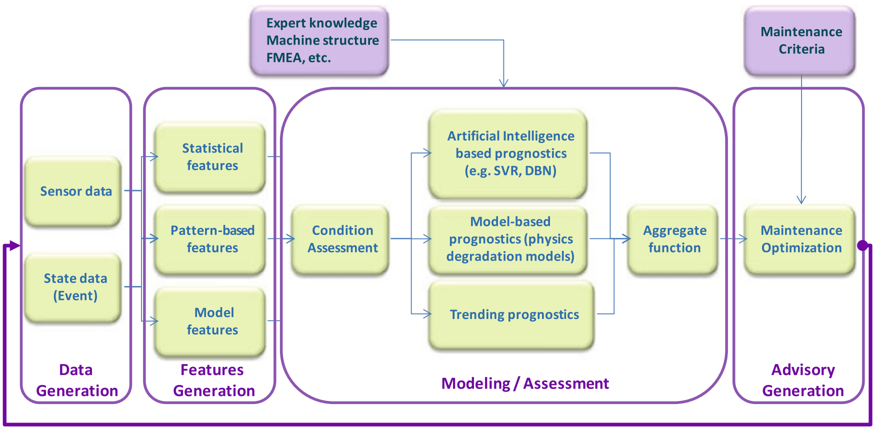


Fig. 8.3 Overall architecture of POM CBM framework

8.4 Cost Model for Optimal Maintenance Planning

8.4.1 Maintenance Cost Versus Product Quality Degradation

For both case studies a cost model considering the trade-off between the cost of maintenance actions and the cost of quality degradation is built. The cost function is continuously updated as new information about the condition and performance of the equipment becomes available from the monitoring system. This maintenance cost information enables optimal maintenance planning based on the real performance and degradation of the considered components or systems.

Figure 8.4 illustrates the advantage of using predictive information to schedule preventive maintenance actions compared to time-based preventive maintenance scheduling. The predictive information takes into account the changing usage rates and environmental conditions which influence the degradation process, while the time-based maintenance actions assume a fixed degradation over time. The timing of preventive maintenance actions are plotted on the x-axis, where tM is the time of maintenance based on monitoring information and tP is the time of the time-based preventive maintenance action. On the y-axis is the decision rule based on the monitored feature (β). This decision rule can have different implementations like, for example, a fixed threshold on a condition monitored parameter. For the first case study in this paper a decision rule based on the maximal profit is implemented, this will be discussed in Sect. 8.4.2. The predictive maintenance policy prevents, for the first maintenance action, the loss due to quality degradation of the product by performing maintenance earlier compared to the time-based policy as depicted in Fig. 8.4. For the second maintenance action the quality degradation is less than anticipated by the time-based policy, a loss due to too much maintenance is incurred here compared to a predictive policy. This shows that a trade-off between the cost of quality degradation and the cost of maintenance should be used in an optimization process to come to an optimal maintenance policy. This is illustrated, together with the ability of the predictive maintenance policy to incorporate the changing quality degradation by changing usage rates and environmental conditions, in the next sections of this paper.

8.4.2 1st Case Study: Seal Quality Monitoring in a Packing Machine

For the first case study, the relevant feature, which is correlated with the performance of the machine, is the percentage of bad bags produced by the sealing machine. Zero padding technique is applied to this feature in order to avoid divergences due to bad seals in beginning of production. This feature is used to represent quality degradation of the produced bags in a cost function as follows:

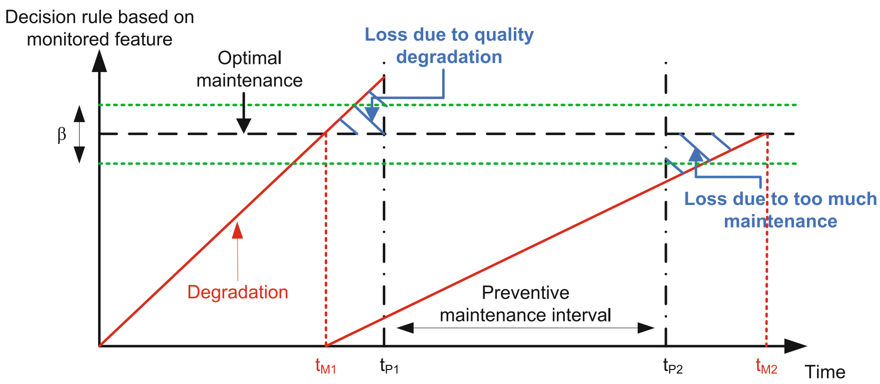


Fig. 8.4 Advantage of a predictive maintenance policy over a time-based preventive maintenance policy considering a trade-off between the cost of maintenance actions and the cost of quality degradation for different deterioration rates

$$P_t = (P \times (1 - \alpha)) - (C \times \alpha) - (M/n), \tag{8.1}$$

where:

t	Time after previous maintenance action
P_t	Profit per bag (€) until time t
P	Profit for one good sealed bag (€)
C	Cost for one bad sealed bag (€)
M	Maintenance cost
n	Number of produced bags until time t
α	Percentage of bad sealed bags until time t

Based on this cost function it is possible to come up with a decision rule, which determines when maintenance should be performed. For this specific case study maintenance is performed when:

$$P_t < P\text{max} \times (1 - \beta), \tag{8.2}$$

where:

t	Time after previous maintenance action
P_t	Profit per bag (€) until time t
P_{max}	Maximal profit per bag (€) until time t
β	Maintenance percentage

This means that at each time t, when a bag is produced, the profit per bag P_t is updated according to the new information on the percentage of bad sealed bags α. When P_t becomes smaller than a certain percentage, which is determined by parameter β, of the maximal profit per bag Pmax until t, a preventive maintenance action should be performed. The reason why P_t is allowed to decrease compared to Pmax is because the considered feature of percentage of bad sealed bags is not monotonically increasing and may fluctuate due to 'self cleaning' phenomenon. This phenomenon consists of disappearance of dirt during production process after it gets accumulated in the sealing jaws. The maintenance percentage β is the parameter in the decision rule, which determines when maintenance should be performed in order to maximize the profit per bag P_t. The determination of the optimal value of β is performed based on data coming from experiments on the packaging machine itself. Simulations on this data were used to determine the value of β which optimizes P_t. Fig. 8.5 shows the result of this optimization where $P = 10€$, $C = 10€$. The maintenance cost M for the specific case study performed in this paper is 200€, which according to the optimization means that a value of 0.02 for the maintenance percentage β is optimal. Of course it is possible to update the value of β continuously when more data becomes available.

In order to quantify the added value of decision making based on real-time evaluation of the earlier introduced cost function and decision rule, a comparison is made between this policy and the maintenance actions that were performed in real life. The maintenance actions performed in real life are based on the experience of the operator. When the operator believes that the packaging machine produces too many bad sealed bags, a preventive maintenance action is performed. The results of the comparison are given in Fig. 8.6, which shows the percentage of bad bags (α) and profit per bag Pt in function of the number of produced bags. Both a reference maintenance scenario, which shows the real life situation, as well as a predictive maintenance policy are presented.

The maintenance timing reference that is shown in Fig. 8.6 is the timing of the maintenance actions performed based on the experience of the operator without making use of the monitored feature (α) for decision making. A total profit per bag $P_{TOT,\ REF}$, which is the profit of the reference scenario for the entire experiment, is also calculated in order to make comparison with the predictive maintenance

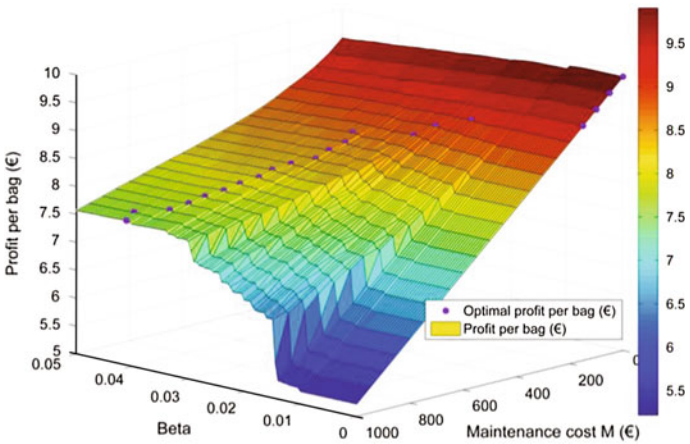


Fig. 8.5 Determination of β which maximizes the profit per bag P_t

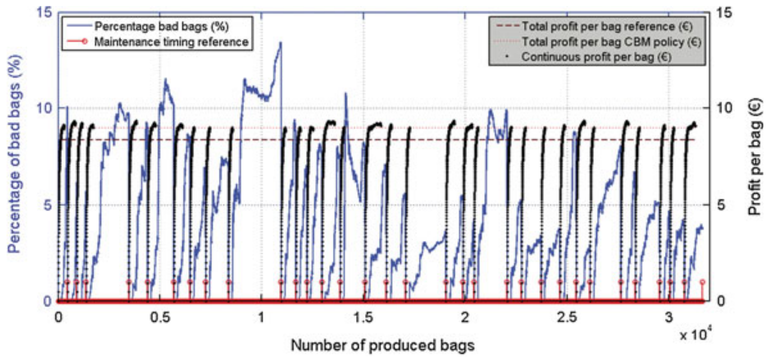


Fig. 8.6 Comparison between reference maintenance policy and predictive maintenance policy

policy possible. By using the monitored feature, percentage of bad bags (α), it is possible to implement the predictive maintenance policy to the real life data collected from the packaging machine. The continuously updated profit per bag Pt, which is used to schedule maintenance as described before, is presented in Fig. 8.6. Based on the decision rule (Eq. 8.2) a preventive maintenance action is scheduled which takes into account the trade-off between the cost of quality degradation and the cost of maintenance. From this simulation it is clear that in general the operator waited too long to perform a preventive maintenance action, which results in a decrease in profit per bag due to quality degradation of the produced bags. A total profit per bag for the entire experiment is calculated for both the reference scenario ($P_{TOT,REF} = 8.3436$€) and the predictive maintenance scenario ($P_{TOT,PdM} = 8.9720$€). An increase of 7.01 % in the total profit is possible by implementing a predictive maintenance policy. That incorporates the changing quality degradation due to different usage rates and environmental conditions. This predictive maintenance policy makes it possible to monitor the profit per bag in real-time, which assists the operator to perform maintenance at the optimal time.

8.4.3 2nd Case Study: Print Quality Monitoring in Copiers

For the second case study, the monitored feature is directly correlated to the quality degradation of the product [16]. For a photocopier, quality degradation can be seen as bad copied pages. The general overview of how the predictive information is used to optimally schedule maintenance actions based on a trade-off between maintenance costs and quality degradation costs is shown in Figs. 8.7 and 8.8. Figure 8.7 depicts the prediction of the evolution of the monitored featured and the corresponding degradation. A lower threshold [feature value of 800 (TH1)] and an upper threshold [feature value of 1,500 (TH2)] are used to describe the quality degradation over time. Before reaching the lower threshold no bad copies are produced, which means the photocopier is in perfect working condition. When the degradation feature reaches the lower threshold, quality degradation of the produced copies starts and evolves through time according to a quality degradation function. This quality degradation function is assumed to be linear in this paper and is shown in Fig. 8.8. The quality degradation function describes a linear relation between the monitored feature and the probability of producing bad copies. When the monitored feature reaches the upper threshold, the probability of producing bad copies equals 1, which means only bad copies are produced at this time and the photocopier is in a failed state. The time that the feature value reaches the lower threshold is defined as t_1 and the time of reaching the upper threshold is defined as t_2.

Based on the predicted deterioration and the corresponding quality degradation function it is possible to optimally schedule preventive maintenance actions by optimizing a cost function. Each time new monitoring information and a corresponding prediction about the state of the component becomes available, the cost function and preventive maintenance timing is updated. The cost function is defined as follows:

$$P_t(t_a) = \frac{(t_1 n_1 \times P) + ((1 - D(x)) \times n_\Delta \times P) - (D(x) \times n_\Delta \times C) - M}{n_1 + n}, \quad (8.3)$$

where:

$P(t_a)$	Profit per copy (€) when maintenance is performed at time t_a
t_1	Time when lower threshold of monitored feature is reached
n_1	Number of copies produced until reaching the lower threshold of the monitored feature
n_Δ	Number of copies produced between t_1 and t_a
P	Profit for one good copy (€)
C	Cost for one bad copy (€)
M	Maintenance cost (€)
D(x)	Percentage of bad copies between t_1 and t_a

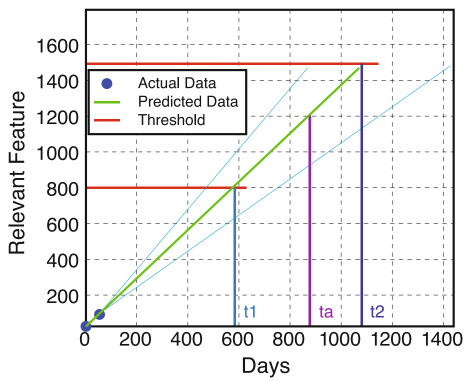


Fig. 8.7 Prediction of deterioration

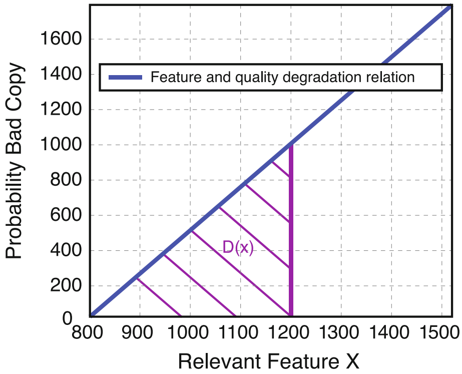


Fig. 8.8 Quality degradation function

In this cost function the quality degradation is incorporated by the function D(x), the percentage of bad copies between t_1 and t_a, which reflects the Fig. 8.8. The function D(x) is calculated as follows:

$$D(x) = \int_{x_{t1}}^{x_{ta}} d(x)dx/(TH_2 - TH_1), \tag{8.4}$$

where:

D(x) Percentage of bad copies between t_1 and t_a
x_{t1} Feature value at time t_1
x_{ta} Feature value at time t_a
d(x) Quality degradation function
TH_1 Feature threshold value where quality degradation starts
TH_2 Feature threshold value where quality degradation is maximal

Based on the deterioration prediction (Fig. 8.7) and the quality degradation function (Fig. 8.8), it is possible to determine the optimal time to perform maintenance by evaluating the cost function defined in Eq. (8.3) for different timings of maintenance. The time where the profit curve is maximized is the optimal time to perform preventive maintenance. This is shown in Figs. 8.9 and 8.10, respectively versus relative days after lower threshold is reached and absolute days. This profit curve, together with the corresponding optimal time to perform preventive maintenance is updated each time new predictive information becomes available.

Fig. 8.9 Profit curve for relative time unit since lowest threshold

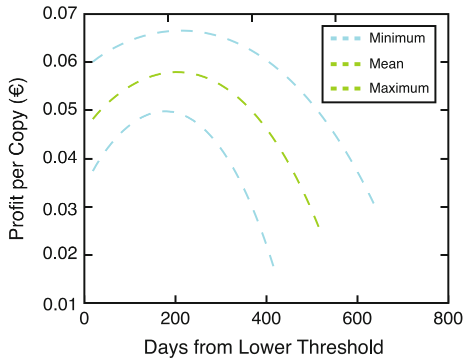
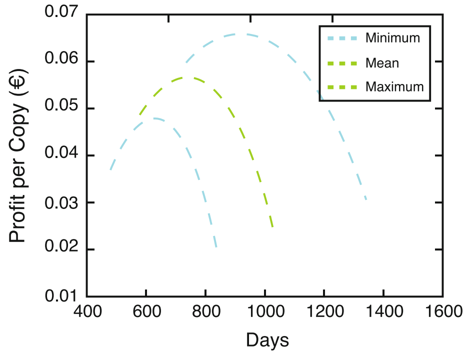


Fig. 8.10 Profit curve for absolute time unit



8.5 Conclusions

Condition-Based Maintenance (CBM) optimization with regard to maintenance cost and quality degradation has been theoretically described in this paper and experimentally validated on two industrial use cases.

This CBM optimization utility which fits in the POM CBM framework for designing and implementing CBM for industrial cases, proved to be able to deal with real data from studied use cases, where complexities in predictive features, like local minima, measurement noise, and abrupt usage changes could take place.

Based on the developed profit maximization technique, it is possible to optimize robustly the maintenance in real time by monitoring online degradation of the product.

Regarding the trade-off between maintenance cost and quality degradation cost, the added value of predictive information in maintenance optimization is substantial.

In future research, we will cover complexities met in industrial cases, like 'self cleaning' phenomenon described earlier in this paper resulting on fluctuating predictive features, and automatic updating of optimization parameters (β in Fig. 8.5).

The robustness of the optimization utility will also be validated with more industrial cases and more online dataset.

Acknowledgments This work has been carried out within the framework of the Prognostics for Optimal Maintenance (POM) project (grant nr. 090045) which is financially supported by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). POM (www.pom-sbo.org) is a cooperation of the project partners Flanders's Mechatronics Technology Centre (FMTC), Centrum voor Industrieel Beleid (CIB, K. U. Leuven), Interdisciplinary Institute of Broadband Technology (IBBT-ETRO), Department of Production engineering, Machine Design and Automation (PMA, K. U. Leuven), Department of Applied Engineering in De Nayer Institute (DNI), Department of Mechatronics, Biostatistics and Sensors (MeBios, K. U. Leuven), Department of Declarative Languages and Artificial Intelligence (DTAI, K. U. Leuven) and Department of Signals, Identification, System Theory and Automation (SCD, K. U. Leuven). The authors wish to thank all the POM project partners for their valuable advices.

References

1. Prognostics for Optimal Maintenance (POM) project (www.pom-sbo.org)
2. Bey-Temsamani A, Bartic A, Vandenplas S (2011) Prognostics for optimal maintenance (POM): An integrated solution from data capturing to maintenance decision. In: Proceedings of the 24th international congress on condition monitoring and diagnostics engineering management (COMADEM). Stavanger, Norway, pp 370–379
3. Sharma A, Yadava GS, Deshmukh SG (2011) A literature review and future perspectives on maintenance optimization. J Qual Maintenance Eng 17(1):5–25
4. Welte M, Vatn J, Heggset J (2006) Markov state model for optimization of maintenance and renewal of hydro power components. In: Paper presented at 9th international conference on probabilistic methods applied to power systems, Sweden, 11–15 June 2006
5. Changyou L, Minqiang X, Song G, Rixin W (2010) Multi objective maintenance optimization of the continuously monitored deterioration system. J Sys Eng Electron 21(5):791–797
6. Zhigang T, Tongdan J, Bairong W, Fangfang D (2011) Condition based maintenance optimization for wind power generation systems under continuous monitoring. Renewable Energy 36:1502–1509
7. Bouvard K, Artus S, Bérenguer C, Cocquempot V (2011) Condition-based dynamic maintenance operations planning & grouping. Application to commercial heavy vehicles. Reliab Eng Sys Saf 96(6):601–610
8. van der Weide JAM, Pandey MD, van Noortwijk JM (2010) Discounted cost model for condition-based maintenance optimization. Reliab Eng Sys Saf 95(3):236–246
9. Van Horenbeek A, Pintelon L, Muchiri P (2010) Maintenance optimization models and criteria. Int J Sys Assur Eng Manage 1(3):189–200
10. Muller A, Crespo A, Iung B (2008) On the concept of e-maintenance: review and current research. J Relib Eng Syst Saf 93:1165–1187
11. Sheppard, Kaufman M, Wilmering T (2008) IEEE standards for prognostics and health management. In: Proceedings of IEEE autotestcon. Salt Lake City, Utah, USA, pp 97–103
12. De Ketelaere B, De Baerdemaeker J, Kamers B (2004) Sealing process inspection device. Patent No WO2004099751, 18 Nov 2004
13. Ostyn B, Darius P, De Baerdemaeker J, De Ketelaere B (2007) Statistical monitoring of a sealing process by means of multivariate accelerometer data. J Qual Eng 19:299–310

14. Bey-Temsamani A, Engels M, Motten A, Vandenplas S, Ompusunggu A (2009) A Practical approach to combine data mining and prognostics for improved predictive maintenance. In: Proceedings of the 15th ACM SIGKDD conference on knowledge discovery and data mining. Paris, France, pp 37–44
15. Bey-Temsamani A, Engels M, Motten A, Vandenplas S, Ompusunggu A (2009) Condition-based maintenance for OEM's by application of data mining and prediction techniques. In: Proceedings of the 4th world congress on engineering asset management (WCEAM). Athens, Greece, pp 543–551
16. Tse MK (1998) Advanced computer-controlled instrumentation for electrophotography. In: SEPJ 40th anniversary Pan-Pacific imaging conference. Tokyo, Japan, 15–17 Jul 1998

Chapter 9
Optimizing Preventive Maintenance Strategies for Linear Assets

Y. Sun, C. Fidge and L. Ma

Abstract Linear assets are engineering infrastructure, such as pipelines, railway lines, and electricity cables, which span long distances and can be divided into different segments. Optimal management of such assets is critical for asset owners as they normally involve significant capital investment. Currently, Time Based Preventive Maintenance (TBPM) strategies are commonly used in industry to improve the reliability of such assets, as they are easy to implement compared with reliability or risk-based preventive maintenance strategies. Linear assets are normally of large scale and thus their preventive maintenance is costly. Their owners and maintainers are always seeking to optimize their TBPM outcomes in terms of minimizing total expected costs over a long term involving multiple maintenance cycles. These costs include repair costs, preventive maintenance costs, and production losses. A TBPM strategy defines when Preventive Maintenance (PM) starts, how frequently the PM is conducted and which segments of a linear asset are operated on in each PM action. A number of factors such as required minimal mission time, customer satisfaction, human resources, and acceptable risk levels need to be considered when planning such a strategy. However, in current practice, TBPM decisions are often made based on decision makers' expertise or industrial historical practice, and lack a systematic analysis of the effects of these factors. To address this issue, here we investigate the characteristics of TBPM of linear assets, and develop an effective multiple criteria decision making approach for determining an optimal TBPM strategy. We develop a recursive optimization

Y. Sun (✉) · C. Fidge · L. Ma
CRC for Integrated Engineering Asset Management, Brisbane 4001 QLD, Australia
e-mail: y3.sun@qut.edu.au

Y. Sun · L. Ma
School of Engineering Systems, Queensland University of Technology, Brisbane 4001 QLD, Australia

C. Fidge
Faculty of Science and Technology, Queensland University of Technology, Brisbane 4001 QLD, Australia

J. Lee et al. (eds.), *Engineering Asset Management 2011*, Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_9,

equation which makes it possible to evaluate the effect of different maintenance options for linear assets, such as the best partitioning of the asset into segments and the maintenance cost per segment.

9.1 Introduction

Engineering assets can be divided into three categories:

1. Non-linear assets. These assets can be further classified as component-based assets such as machines; mobile assets such as vehicles; and fixed physical assets such as facilities [1].
2. Linear assets. These assets include engineering structures or infrastructure that usually span long distances and can be divided into different segments that provide the same function but are subjected to different operating loads and environmental conditions. Pipelines, cables, roads, and railway lines are examples of linear assets.
3. Hybrid systems. These systems comprise both linear and non-linear assets. An example is a water supply network that consists of pipelines and pump stations.

In this paper, we focus on linear assets as they are usually of large scale, play an important socioeconomic role in our society and have a number of distinguishing features [2]. One such feature is that linear assets are physically presented in the form of lines. Different lines may connect together to form a network. Linear assets often involve a large amount of capital investment, and it is therefore imperative to manage them effectively. Currently, Preventive Maintenance (PM) is commonly applied to improve the reliability of linear assets and reduce sudden failure risks. Both Reliability (or Risk) Based PM (RBPM) and Time Based PM (TBPM) strategies are implemented in modern industries. In previous work [3], we discussed the optimization of RBPM strategies. Here we instead study TBPM strategy optimization problems. Furthermore, we limit our study to individual lines of a linear asset, rather than a network which consists of multiple lines.

A major task for planning TBPM strategies is to find the best answer to the following questions:

1. When should we start PM for a specific asset?
2. What is the best PM interval?
3. In each PM action, which segments, identified by their locations and lengths, should be preventively maintained?

For the first question, research has shown that PM should not be implemented before the wear-out starting time as PM for assets with exponentially distributed failures is not effective [4]. However, the best time to start preventive maintenance for an asset in the wear-out stage (i.e., with an increasing failure rate) still needs to be determined. In current practice, some companies simply start their PM at the wear-out points of their assets. Our analysis below shows this PM lead time is often too early, resulting in low benefits in terms of reliability improvement per unit expenditure. On the other hand, starting PM too late will also reduce cost efficiency, and result in high failure risks.

The second question is a standard problem in PM decisions.

The third question is raised because a linear asset often spans a long distance. In this case, its segments at different locations may be subject to different environmental conditions and have different failure rates. Moreover, different segments may have different ages. As a result, the contribution of different segments to the whole system's reliability can be different. For each PM action, we need to repair or replace those segments whose improvement produces the highest benefit to system reliability. Apart from determining the PM location, we also need to decide the length of segments for each PM action. As linear assets span long distances, we usually have to preventively maintain a part, instead of the whole, of an asset. Note that in most circumstances, these questions are interrelated and thus must be analyzed together to determine an optimal PM strategy. In addition, various constraints such as mission time and minimal reliability requirements also need to be considered.

In current practice, TBPM decisions are often made based on decision makers' expertise or industrial historical practice. These decisions are often far away from optimal due to a lack of systematic analysis of the effects of these factors. An effective method for optimizing TBPM strategies of linear assets has yet to be developed. In this paper, we develop an interactive Split System Approach (SSA) [5] based approach to address this issue.

The rest of the paper is organized as follows: Section 9.2 presents a brief literature review; A TBPM optimization approach is presented in Sect. 9.3; followed by a numerical example in Sect. 9.4. Conclusions are given in Sect. 9.5.

9.2 Literature Review

Research on linear asset management has attracted considerable attention from engineering asset management practitioners and academic researchers, although previous research has focused on specific types of linear assets. For example, Jones and McManus [6] conducted a life-cycle analysis of five different 11 kV electrical power cables. Brito and Almeida [7] developed a risk ranking model for natural gas pipelines based on multi-attribute utility theory. Their model can incorporate the preferences and behaviors of decision makers quantitatively. The pipelines were divided into a number of smaller sections to model the variable conditions, such as the failure rate, of pipeline segments along a route. On the other hand, Ahern and Anandarajah [8] developed a ranking model for prioritizing new railway projects using weighted integer goal-programming. Their model considers the entire railway associated with the projects. Castanier and Rausand [9] developed a preventive maintenance optimization approach for submarine oil pipelines exposed to internal corrosion-erosion based on the classical PF interval model. However, they considered only a single failure of one pipe segment at a time. Similarly, Creig-Smith [10] used both destructive and non-destructive test methods to determine the corrosion conditions of water steel pipeline segments for renewal decision support. Li et al. [11] developed a management information system for mining railway transportation equipment without considering railway track management issues. To address sparse failure issues, Cagno et al. [12] conducted a case study to estimate the prior distributions of gas pipeline failures using the Analytical Hierarchy Process (AHP) and a Bayesian approach. Coffen-Smout and Herbert [13] presented an overview of submarine cable management challenges, while Link [14] studied the renewal costs of motorways in Germany using a econometric analysis. In Link's study, the road's condition was not explicitly considered.

While all of these research results are useful for maintenance decision support, conventional engineering asset information management systems or decision support tools are not suitable for linear assets because of their hierarchical asset structure, which uses parent–child relationships to link system, assembly, component, and part-type hierarchies [15]. To address linear asset information management issues, Maximo developed a tool called the Linear Asset Manager [16]. However, this tool does not have a decision analysis function. To support renewal decisions of linear assets, we previously [2] developed a decision making process via studying the characteristics of linear assets. Here we extend our previous work to consider time based preventive maintenance of linear assets.

9.3 Methodology Development

A number of factors can affect a preventive maintenance strategy. These include asset maintenance costs, asset heath conditions, asset capacity demands, asset operational time demands, maintenance performance, human resource availability, and logistics, as well as legal, environmental, and social requirements. A common objective for deciding PM strategies is to minimize the total asset management cost. As a result, other factors become constraints that will limit the capability of an organization to achieve this objective. In previous work [3], we analyzed these constraints, and showed that the required minimum mission time and the minimum acceptable mission reliability are the two most critical constraints. The required minimum mission time is the required minimum time for an asset to implement a particular task without interruptions. The minimum acceptable mission reliability is the minimum reliability required by a particular asset to implement a particular task. In this paper, we focus on optimizing a TBPM strategy through minimizing its associated expected total cost under these two critical constraints.

The expected total cost includes the anticipated repair cost and the expected preventive maintenance cost. The repair cost is the expense caused by the failure of an asset, including the production loss, due to repairs. The PM cost is the cost due to conducting PM activities, including production losses caused by preventive maintenance. The modeling approach for optimizing reliability based preventive maintenance strategies [3] can be largely applied to time based PM strategies. However, in RBPM decision making, only a single variable (reliability or risk threshold) needs to be optimized, whereas in TBPM decision making at least two variables need to be optimized simultaneously. These two variables are the PM start time and the PM interval. Therefore, TBPM strategy optimization is a multi-objective optimization problem. For linear assets, we also need to consider the optimal segmentation of the asset. For a long linear asset, replacing the entire asset in every PM action is usually unnecessary and too costly. In practice, a long linear asset is generally divided into a number of segments and only one segment is preventively repaired in each PM action. In this case, we need to optimize the partitioning into segments as well.

Let $R_{\min}$ denote the minimum acceptable mission reliability; $T_{\min}$ denote the required minimum mission time; C_T denote the expected total cost; T_1 denote the PM start time; R_0 denote the conditional reliability, M denote the number of linear asset segments; $\vec{T}_p$ denote the PM interval; T_{pi} denote the interval between the $(i-1)$th PM action and the ith PM action $\left(i = 2, 3, \ldots, N\left(T, M, T_1, \vec{T}_p\right)\right)$; and $N\left(T, M, T_1, \vec{T}_p\right)$ denote the required number of PM actions over the period of a given decision horizon T. As preventive maintenance should be implemented during the wear-out stage, for simplicity, we start to measure T from the wear-out starting point. If $C_R\left(T, M, T_1, \vec{T}_p\right)$ denotes the expected repair cost and $C_M\left(T, M, T_1, \vec{T}_p\right)$ denotes the expected PM related cost, the optimization problem considered in this paper can be defined as follows:

$$C_T = \min\left[C_R\left(T, M, T_1, \vec{T}_p\right) + C_M\left(T, M, T_1, \vec{T}_p\right)\right], \tag{9.1}$$

$$\begin{aligned} \text{s.t.} \quad & T_1 \geq T_{\min}, \\ & T_{pi} \geq T_{\min}, \quad i = 2, 3, \ldots, N\left(T, M, T_1, \vec{T}_p\right), \\ & R_0 \geq R_{\min}. \end{aligned}$$

The expected repair cost increases with the number of failures while the expected PM cost increases with the number of PM actions. From our previous study [17], these two costs are given by

$$C_R\left(T, M, T_1, \vec{T}_p\right) = k_R \times \left[1 - R\left(T, M, T_1, \vec{T}_p\right)\right] \times 100, \tag{9.2}$$

and

$$C_M\left(T, M, T_1, \vec{T}_p\right) = k_M \times N\left(T, M, T_1, \vec{T}_p\right), \tag{9.3}$$

where parameters k_R and k_M are the corrective repair cost rate with a unit of \$ per percentage of failures, and PM cost rate with a unit of \$ per PM action, respectively. Function $R\left(T, M, T_1, \vec{T}_p\right)$ is the overall reliability of the asset at time T.

In general, $N\left(T, M, T_1, \vec{T}_p\right)$ can be calculated using the following equation:

$$\sum_{i=2}^{N\left(T,M,T_1,\vec{T}_p\right)} T_{pi} \leq (T - T_1) < \sum_{i=2}^{N\left(T,M,T_1,\vec{T}_p\right)+1} T_{pi}, \tag{9.4}$$

In reality, for management simplicity, an equidistant PM interval T_p is often adopted. Let $\lfloor x \rfloor$denote the 'floor' function applied to x, i.e., the largest integer less than or equal to x. In this case, $N\left(T, M, T_1, \overleftarrow{T}_p\right)$ is given by

$$N(T, M, T_1, \vec{T}_p) = \left\lfloor \frac{T - T_1}{T_p} \right\rfloor, \tag{9.5}$$

$R\left(T, M, T_1, \vec{T}_p\right)$ can be calculated using our split system approach [5]. Assume that a linear asset is virtually divided into M segments with no overlapping, and m segments are repaired in j PM cycles (actions) ($j = 0, 1, 2, 3, \ldots, N\left(T, M, T_1, \vec{T}_p\right)$). Let L_k indicate that segment k ($k \leq m$) receives its last repair in the L_kth PM action ($L_k \leq j$). The reliability of each segment after a PM action is $R_i(\tau)_0$. Parameter τ is a relative time scale which is reset to be zero after each PM action. Define $T_{p1} = T_1$. Then the conditional reliability function of a linear asset after the jth PM cycle is given by

$$R_{sc}(\tau)_j = \frac{R_s\left(\tau + \sum_{i=1}^{j} T_{pi}\right)_0 \times \prod_{k=1}^{m} R_k\left(\tau + \sum_{i=L_k+1}^{j} T_{pi}\right)_{L_k}}{\prod_{k=1}^{m} R_k\left(\tau + \sum_{i=1}^{j} T_{pi}\right)_0}, \tag{9.6}$$
$$\left[0 \leq \tau \leq T_{p(j+1)}, \quad j = 0, 1, 2, \ldots, N\left(T, M, T_1, \vec{T}_p\right)\right],$$

In Eq. (9.6), let $\sum_{i=1}^{j} T_{pi} = 0$ when $j < 1$, and $\sum_{i=L_k+1}^{j} T_{pi} = 0$ when $L_k + 1 > j$. The reliability of the asset can be calculated using a heuristic approach. Function $R_{sc}(\tau)_j$ is the conditional reliability of the linear asset after the jth PM action. Function $R_s(\bullet)_0$ is the overall reliability of the asset before any repairs. Functions $R_k(\bullet)_0$ and $R_k(\bullet)_{L_k}$ return the reliability of segment k before any repairs and after the L_kth PM action, respectively. Equation (9.6) needs to be calculated recursively, so a computer program is normally needed to support the calculation. After the conditional reliability of an asset has been calculated, its overall reliability can be calculated using the following recursive equation:

$$R(\tau|T, M, T_1, \vec{T}_p)_j = R_k\left(\sum T_{pi}\right)_{j-1} R_{sc}(\tau)_j,$$
$$\left[0 \le \tau \le T_{p(j+1)}, \quad j = 0, 1, 2, \ldots, N\left(T, M, T_1, \vec{T}_p\right)\right], \tag{9.7}$$

where $R_k\left(\sum T_{pi}\right)_{j-1}$ is the reliability of segment k being repaired at the jth PM action. Function $R\left(\tau|T, M, T_1, \vec{T}_p\right)_j$ is the overall reliability of the asset after the jth PM action.

9.4 A Numerical Example

Here we use a numerical example to demonstrate the applicability of our model. In this example, a pipeline is assumed to be 100 m long, with the following reliability function,

$$R_s(t)_0 = \exp\left[-\left(\frac{t}{60}\right)^{2.3}\right], \tag{9.8}$$

where $R_s(t)_0$ is the reliability function of the pipeline.

As the failures of the pipeline are assumed to follow a two parameter Weibull distribution, the owner of the pipeline decides to implement a TBPM strategy to minimize the total expected cost and control the risk of sudden failure. For simplicity, the pipeline is further assumed to be homogeneous. It can be divided into M equal segments for preventive maintenance. According to this assumption, the reliability of each segment before Preventive Maintenance, $R_k(t)_0$, is

$$R_k(t)_0 = \sqrt[M]{\exp\left[-\left(\frac{t}{60}\right)^{2.3}\right]}, \quad k = 1, 2, 3, \ldots, M. \tag{9.9}$$

It is also assumed that only one segment is preventively maintained in each PM action. The reliability of a segment after a PM action follows the same reliability function, i.e., Equation (9.9), but the age is reset to zero due to the improving effect of the PM action. Note that in general the pipeline's condition after a PM action is improved but cannot be restored to "as good as new". When the preventively maintained segments are very short, i.e., the pipeline is divided into many segments, the condition of the pipeline after a PM action can be assumed to be as "as bad as old".

The Preventive Maintenance cost rate is assumed to be \$510 per meter per PM action, and the corrective repair cost rate is assumed to be \$4,000 per percentage of failure probability. The decision horizon is assumed to be 40 years. The owner of the pipeline needs to determine the optimal segmentation of the pipeline, the optimal PM start time, and the optimal PM interval.

A series of scenarios were analyzed. Some results are illustrated in Figs. 9.1, 9.2, 9.3, 9.4, 9.5. In these figures, three dimensional figure a present the total

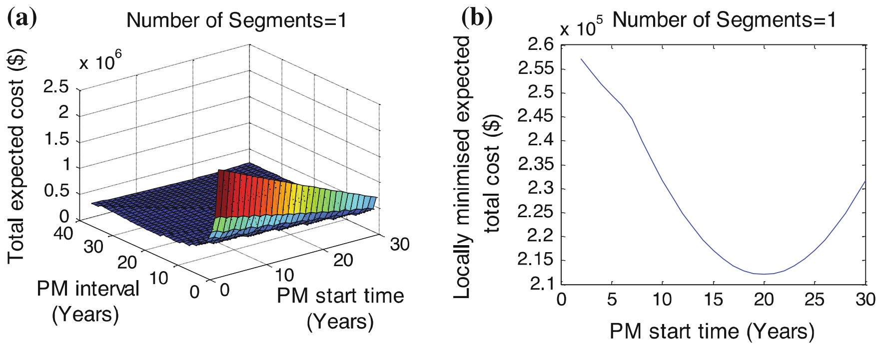


Fig. 9.1 The changes to the total expected cost (**a**) and the locally minimized expected total cost (**b**) when $M = 1$

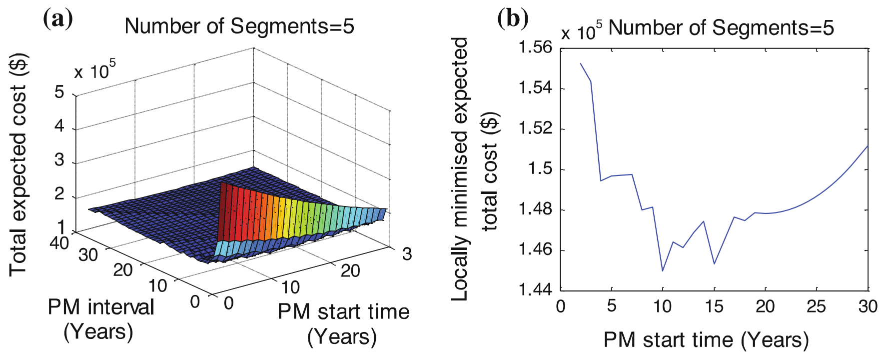


Fig. 9.2 The changes to the total expected cost (**a**) and the locally minimized expected total cost (**b**) when $M = 5$

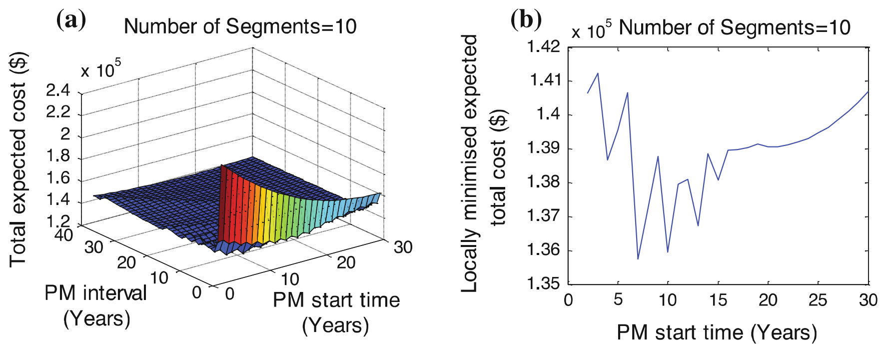


Fig. 9.3 The changes to the total expected cost (**a**) and the locally minimized expected total cost (**b**) when $M = 10$

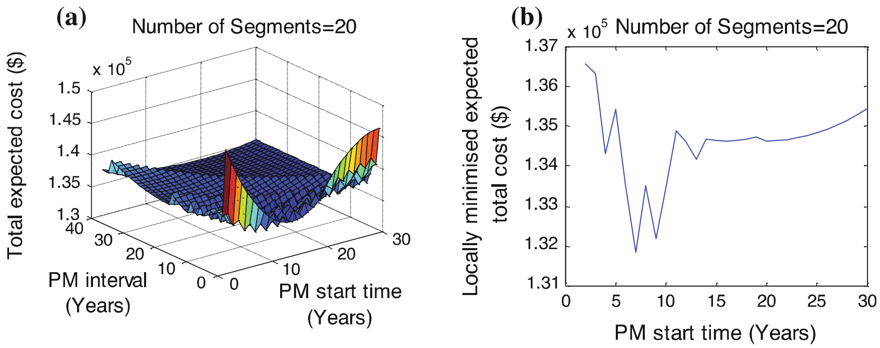


Fig. 9.4 The changes to the total expected cost (**a**) and the locally minimized expected total cost (**b**) when $M = 20$

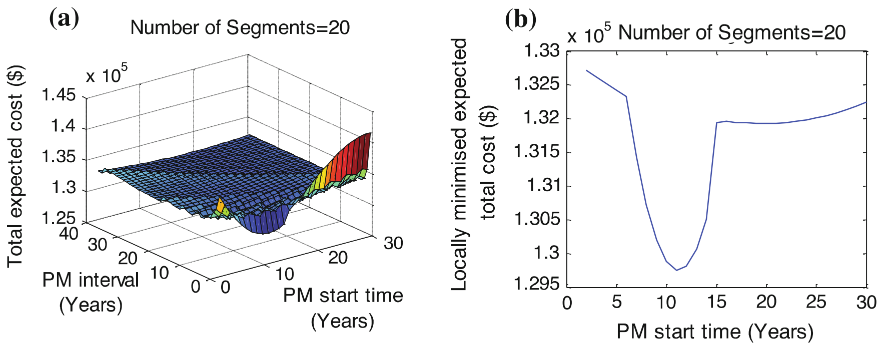


Fig. 9.5 The changes to the total expected cost (**a**) and the locally minimized expected total cost (**b**) when $M = 50$

expected costs, which include the corrective repair costs and PM costs corresponding to different partitionings, PM start times and intervals. From these three dimensional figures, it can be seen that for different segmentations, the total expected costs have different presentations. When the whole pipeline is maintained as a single segment, the total expected cost changes with the PM start time and intervals smoothly, but when the pipeline is divided into multiple segments for PM, the total expected cost changes with the PM start time and intervals alternately. However, in all these figures, the cost values near the point (0, 0) are much higher than in other areas. This means that too many PM actions normally result in a high cost. Although there is a lowest point in each three dimensional figure, which corresponds to the lowest total expected cost, it is hard to identify these points in figures as displayed here, due to there being so many nearby values.

To address this issue, the two dimensional figure b in Figs. 9.1, 9.2, 9.3, 9.4, 9.5 present the locally minimized expected total cost, which is the minimal total expected cost corresponding to a specific number of segments and a specific PM

Table 9.1 The optimal PM start time and interval with different numbers of segments

Number of segments	Optimal starting time (years)	Optimal PM interval (years)	Minimal total cost over 40 years ($)	Minimum mission reliability
1	20	20	212,080	0.9232
5	10	5	144,950	0.9466
10	7	3	135,750	0.9305
20	7	1.5	131,800	0.9355
50	11	0.5	129,750	0.9591

start time. This information highlights the globally optimized PM start time more clearly. From these two dimensional figures, it can be seen that all the locally minimized expected total cost curves have the lowest points, although some of them fluctuate with the PM start time. It can also be seen that the lowest values corresponding to different segmentations are different. This information can be used to decide on the optimal segmentation. Nevertheless, when choosing the most suitable segmentation, PM start time and interval, one also needs to consider the conditional reliability.

Summarizing these results, the optimal PM start times, optimal PM intervals, minimal total costs, and minimum mission reliabilities for the various selected maintenance segments are shown in Table 9.1.

If the owner only considers these five options, without considering constraints on minimum mission reliability and mission time, the optimal decisions in terms of the minimum expected total cost is dividing the pipeline into 50 segments, i.e., replacing two meters of pipe every time, starting PM at 11 years and then preventively replacing a segment in every half year. However, conducting a PM action every half a year could be too frequent for the owner because of the frequent service interruptions. If the business demands that the minimal mission time should be greater than 1 year, then dividing the pipeline into 20 segments is the optimal option. If the owner also wants to control sudden failure risks during the mission period and requests that the minimum mission reliability must be higher than 0.94, then dividing the pipeline into five segments is the option. Note that, in this case, the expected total cost will increase by 11.71 %, compared with the optimal option without considering constraints.

9.5 Conclusions

Linear assets are engineering structures or infrastructure that often cross a long distance and can be divided into different segments that provide the same function but under different conditions. Compared with non-linear assets, linear assets normally do not have a clear physical boundary and their segments are often virtually defined based on maintenance activities. Linear assets are crucial to the community and the economy, often involving a large amount of capital investment. They thus need to be optimally maintained. As failures of these assets are often expensive, preventive maintenance is widely applied to them to reduce their sudden failure risks. However, preventive maintenance is also costly. Therefore, finding optimal preventive maintenance strategies for linear assets, especially over a long term, with multiple maintenance cycles, is of strategic importance to linear asset owners.

Here we studied optimization of the Time Based Preventive Maintenance (TBPM) strategy for linear assets, as TBPM is one of the commonly used strategies for improving the reliability of assets in practice. In TBPM, a preventive maintenance action is conducted on a predefined segment based on a prescheduled time. The objective of optimizing a TBPM strategy for a linear asset is to find the optimal number of PM segments, the optimal PM start time, and the optimal PM intervals, to minimize the expected total costs including the repair cost, and preventive maintenance cost. A number of factors such as the required minimal mission time and minimum acceptable mission reliability can limit the ability of an organization to achieve this objective. To produce an effective method for making optimal TBPM decisions we developed here a multi-objective optimization approach that works within multiple constraints. Although the approach was developed for linear assets, it can be extended to non-linear assets straightforwardly.

Acknowledgments This research was conducted within the *CRC for Infrastructure and Engineering Asset Management*, established and supported under the Australian Government's Cooperative Research Centres Program.

References

1. IBM (2008) IBM Maximo Linear Asset Manager 7.1. Available from http://www.ibm.com/podcasts/software/tivoli/maximo/Linear_Asset_mgr6_30.pdf. Accessed 25 June 2010
2. Sun Y et al (2010) Renewal decision support for linear assets. In: The 5th world congress on engineering asset management (WCEAM & AGIC 2010). Springer-Verlag, Brisbane, Australia
3. Sun Y et al (2009) Optimization of the reliability based preventive maintenance strategy. In: The 4th world congress on engineering asset management (WCEAM2009). Springer-Verlag, Athens, Greece
4. Elsayed EA (1996) Reliability engineering. Addison Wesley Longman, Inc. Reading, Massachusetts, pp 378–395
5. Sun Y, Ma L, Mathew J (2007) Prediction of system reliability for multiple component repairs. In: Proceedings of the 2007 IEEE international conference on industrial engineering and engineering management. IEEE, Singapore
6. Jones CI, McManus MC (2010) Life-cycle assessment of 11 kV electrical overhead lines and underground cables. J Cleaner Prod 18(14):1464–1477
7. Brito AJ, de Almeida AT (2009) Multi-attribute risk assessment for risk ranking of natural gas pipelines. Reliab Eng Syst Saf 94(2):187–198
8. Ahern A, Anandarajah G (2007) Railway projects prioritisation for investment: application of goal programming. Transp Policy 14(1):70–80
9. Castanier B, Rausand M (2006) Maintenance optimization for subsea oil pipelines. Int J Press Vessels Pip (The 16th European Safety and Reliability Conference) 83(4):236–243

10. Creig-Smith S (2001) To refurbish or replace steel water pipelines, that is the question. Eng Fail Anal 8(2):107–112
11. Li MY et al (2008) A management information system for mine railway transportation equipment. J China Univ Min Technol 18(3):373–376
12. Cagno E et al (2000) Using AHP in determining the prior distributions on gas pipeline failures in a robust Bayesian approach. Reliab Eng Syst Saf 67(3):275–284
13. Coffen-Smout S, Herbert GJ (2000) Submarine cables: a challenge for ocean management. Mar Policy 24(6):441–448
14. Link H (2006) An econometric analysis of motorway renewal costs in Germany. Transp Res Part A Policy Pract 40(1):19–34
15. IBM (2009) Linear asset manager. Available from http://www.muwg.org/MUWG%202009%20-20Linear%20Asset%20Manager%20Presentation.ppt#256,1,Linear. Accessed on 23 June 2010
16. IBM (2008) Better manage your linear assets with IBM Maximo linear asset manager. Available from http://www.gocfi.com/downloads/maximo/Maximo%20Linear%20Asset%20Manager.pdf. Accessed on 23 June 2010
17. Sun Y, Ma L, Mathew J (2006) Determination of optimal preventive maintenance strategy for serial production lines. In: The 1st world congress on engineering asset management. Springer-Verlag, Gold Coast, Australia

Chapter 10
Identifying Differences in Safe Roads and Crash Prone Roads Using Clustering Data Mining

D. Emerson, R. Nayak and J. Weligamage

Abstract Road asset managers are overwhelmed with a high volume of raw data which they need to process and utilize in supporting their decision making. This paper presents a method that processes road-crash data of a whole road network and exposes hidden value inherent in the data by deploying the clustering data mining method. The goal of this method is to partition the road network into a set of groups (classes) based on common data and characterize the crash types to produce a crash profile for each cluster. By comparing similar road classes with differing crash types and rates, insight can be gained into these differences that are caused by the particular characteristics of their roads. These differences can be used as evidence in knowledge development and decision support.

10.1 Introduction

Road safety is a major concern worldwide with road crashes costing countries between 1 and 3 % of annual Gross Domestic Product. The World Health Organization (WHO) predicts road traffic crashes emerging as the 3rd leading cause of disease or injury burden [1]. The annual economic cost of road crashes in Australia is conservatively estimated at $18 billion per annum, with crashes having the

D. Emerson (✉) · R. Nayak
Computer Science Discipline, Science and Technology, Queensland University of Technology, Brisbane, QLD, Australia
e-mail: d.emerson@connect.qut.edu.au

R. Nayak
e-mail: r.nayak@qut.edu.au

J. Weligamage
Road Asset Management Branch, Queensland Government Department of Main Roads, Brisbane, QLD, Australia
e-mail: justin.z.weligamage@tmr.qld.gov.au

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_10,

potential for devastating social impact [2]. Worldwide, road authorities follow strict design codes applying known engineering principles within the contexts of safety, cost, driver expectation, and economic and environmental parameters [3]. Research is continually contributing to development of these principles.

Historically, statistical studies have been used to analyze crashes from homogeneous road subsets [4], however, there is a demand for analysis of the whole of the network. The ARRB Group, an Australian road research organization, contends that use of safety risk assessment at the network level (rather than concentrating on single sites) can apply consistency in the approach to road safety across the entire network [5]. Depaire et al. [6] provide the theoretical framework and a case study for use of probability-based clustering data mining for analyzing crashes from heterogeneous road sets [6], but the case study is limited to a set of urban roads.

This paper evaluates an applied data mining approach to using clustering data mining using the probabilistic expectation maximization method [6] on a large, diverse rural–urban network, and shows the value of clustering by presenting the practical decision support outcomes of the case study. The method uses relationships between road-related attributes and crash types to build clusters and document a profile of individual cluster crash characteristics, which subsequently can be used to define the risks faced by motorists as they travel a particular group of road segments. This chapter describes the data mining approach used to develop road clusters by deploying the following road and crash characteristics as input variables: surface measures, surface type, age, wear, damage, road geometry, traffic, and crash type. Clustering, using the selected attributes, develops natural groups in an unsupervised fashion, then assigns instances to those clusters, resulting in grouping of instances according to their attributes. The dimensionality of the attribute value range active in each cluster is reduced, and the clusters make new datasets suitable for statistical analysis [6].

Data mining methods are increasingly in demand in business situations [7] and are being applied in many innovative ways in road crash studies [8, 9]. Anderson [10] recently used clustering in a road crash study to analyze crash hotspots using spatial density. The approach presented in this study which focuses on the whole of the network, classified 4 years of crashes, their 1 km road segments, and the road segment crash counts [11]. These crash, road, traffic, and conditions characteristics provided the input variables for the clustering models. Studies such as Quin [12] demonstrate that road factors influence crashes, with significant factors including speed limit, traffic rates, time of day, volume/capacity ratio, percent of passing lanes, shoulder width, number of intersections and driveways. These characteristics coincide with many of the variables in our dataset.

The incentive to investigate data mining clustering as a method of road and crash benchmarking evolved from our preliminary clustering of road characteristics during an investigation into road segment crash proneness [13]. The study showed that road segment clusters developed from road attributes had a diverse range of crash counts, characterizing each cluster's into a category of very low, moderate, or high crash.

The contribution of this paper is to demonstrate that clustering methods can be successfully applied to whole-of-network road-crash data for the development of useful road cluster crash profiles for use in decision support, and to also provide guidelines for method deployment. The paper describes development and outcomes of the study in the following sections. Section 10.2 describes the data organization and pre-processing, while Sect. 10.3 describes the data mining methodology and method assessment. Section 10.4 discusses the methods used in the case study to analyze, organize, and present information, and demonstrates use of the road segment crash profiles in decision making. The conclusion provides the summary of the outcomes and proposes future directions.

10.2 Data Set Development

The dataset utilized is a four-year list of 2004–2007 crash instances populated with the road data, provided by the Queensland Department of Transport and Main Roads (QDTMR). Of the 42,388 crashes reported, sufficient data were available for inclusion of one-third of the crashes for analysis. The attributes provided the necessary road and traffic information to develop a series of road cluster types with sufficient differentiation to allow comparison between high risk and low risk road clusters and their crash types. Road attributes are modeled in the following logical groups:

Road surface, design, and wear (19 attributes) included *roadway surface with* road surface friction, texture depth, seal age, and seal type; *road surface wear and damage with* rutting and roughness; *roadway design:* with speed limit, traffic, divided road, dual or single carriageway, road type such as highway, urban arterial, etc., carriageway type and the lane count; *roadway features* such as roundabouts, bridges, intersections; traffic control with tights, signs, etc.; *traffic rates*; *geometry with* horizontal and vertical alignment and terrain.

Roadway and cracking (24 attributes): consisting of the attributes above and five cracking attributes.

Roadway and crash details (32 attributes): addition of *crash details* such as crash type, road surface wetness, monthly rainfall, day, time, and lighting.

A standardized copy of the data was generated for modeling. The natural range of the AADT attribute, from 61 to 64,452 vehicles per day, is many magnitudes higher than road friction (F60) ranging from 0.08 to 0.66. Standardized ranges were −1.077 to 4.195 versus −4.462 to 4.185 respectively. Cluster models were developed with both the original non-standardized and processed standardized versions and compared.

10.3 Data Mining Methodology, Results and Discussion

This section examines how the clustering methods were configured and discusses the results of the trials. The clustering application selected was the probabilistic method called *expectation maximization* (EM) [6]. This method allocates instances to clusters on the highest likelihood of an instance being in a given cluster. EM provides a model goodness measure, *log likelihood.* Our study progressed through two stages. In stage 1 we performed clustering benchmarking to get to know how the data would perform with the EM clustering method, and to assess the usefulness of resulting clusters. Stage 2 saw benchmarking of the EM configuration parameters in conjunction with the automated cluster count determination function.

In stage 1 we ran a series of tests with combinations of the road and crash attributes, cluster counts, as well as standardized and non-standardized data. Cluster count was set in the EM application configuration interface, and clusters were created with the individual cluster count ranging between 2 and 16 clusters. The models were evaluated in three ways: by comparing the method's output statistics, analyzing the cluster structure, and comparing how well, for a given trial, the clusters differentiated themselves. Extensive visualization was used to compare the clustered attributes values to evaluate usefulness of a given model in its relative contribution to the road domain knowledge.

Results showed that the model quality statistic *log likelihood* was not very useful determining the best models and that producing a standardized dataset had only marginal benefit. Models from standardized and non-standardized datasets in most cases produced identical clusters, while having wildly different log likelihood values; −22.9 verses −55.5 for standardized versus non-standardized datasets.

While searching for the optimal seed configuration value, we found that the variation of seed affected the standardized or non-standardized datasets differently, producing clusters with identical instance allocation in some cases and different clusters in others. We favored seed values that produced identical clusters. Completion of benchmarking at stage 1 provided a set of known clusters, useful for later comparisons in assessment of procedural accuracy and configuration for model quality and suitability.

In stage 2 we sought the ideal clustering count, however, the problem of the cluster count dependency on the seed value needed resolution. We used the traditional data mining *validation method* of developing a model by training on 66 % of the instances and validating the model by testing on the remaining data subset. While EM has its own internal cross validation method to progressively develop and test the model, the training/testing method was extremely valuable in selecting the best seed. An optimal cluster count and seed value was selected for each of three standard attribute sets, based on seeds where the cluster numbers and instance allocations from both training and testing were identical.

For the *road only* dataset the seed value of 100 produced a cluster model with 12 groups. The *road and cracking* dataset the seed of 75 produced 15 clusters.

With *road, cracking and crash* dataset, the seed of 75 produced 8 clusters. The unexpected reduction in cluster count with the inclusion of crash data is thought to be caused by the similarity in crash outcomes for some clusters causing them to coalescence, but this requires conformation.

Once the prime benchmarked clusters were documented, a range of new clusters were generated and evaluated. Configuring a model for a higher cluster count caused individual clusters to split and produce cluster subgroups. The split was created on a significant attribute and results were found to be useful for examination of topics such as serious crashes and wet crashes.

Using model statistics in comparing the performance of our EM implementation with Depaire's work [6] was difficult because of different measures in use, and also in our study the log likelihood, while indicating cluster compactness did not relate to the actual cluster formations and instance allocations. However, comparing models at the information level showed that our clusters, like Depaire's, were able to strongly differentiate on the available attributes; in our case with road type and function, crash types, road surface wetness, speed limit, lane counts, crash type, and so on.

10.4 Case Study: Method, Results and Discussion

The third stage describes how the clustering models were investigated and presents a subset of results from one of our case studies to illustrate how intelligent partitioning of the data can be used in knowledge management and decision making. As with many road studies, this study is driven by goals of discovering how roads contribute to crashes, and how we can use outcomes to improve road safety. Clustering provides an extremely valuable tool to partition the dataset into instance groups with a strong relationship to the attribute properties, and provided clusters relating to our topics of investigation: wet crash and serious crash. Since this was a preliminary clustering study with this dataset, we were investigating ways to extract the value from the models to contribute to the goals. Our investigations included;

Visualizing cluster distributions geographically to contextualize them spatially, e.g., plotting the crashes by latitude and longitude to show geographic cluster distribution and allow relation to all other temporal-spatial data.

Using the characteristics to name the cluster and develop named cluster groups, e.g., formation of the *city urban and regional roads cluster group* from a number of similar clusters.

Plotting clusters by attributes to gather evidence on the big questions of road crash causation, e.g., by plotting cluster versus seal type and relating the attribute range to the cluster's significant properties such as the relative cluster crash rate.

Examining the built in cluster/attribute cross tabulation reports which document the relative contribution of attribute values to the individual clusters and using these values in developing cluster profiles. Example rural highway clusters were found that have high vehicle collision with stationary objects and high vehicle overturn type of crash. We used these relative crash values to create cluster crash profiles.

Filtering clusters by attribute values of interest and using cluster attribute distributions to describe the entity characteristics, e.g., obtaining the crash instance cluster types for a particular road or road type and using the cluster crash profiles to understand the crash characteristics of the road.

Clusters provided a range of new ways to characterize road types and provided new data and evidence for development and evaluation of the inferences that help make our roads safer.

The following section presents some of the outcomes of our preliminary case study. This clustering method was configured to produce a model extended from the 8 main classes identified as the prime clusters. This model was configured with 14 clusters with the purpose of splitting the default clusters in the search of the topics highlighted above: wet and serious crash.

The model was developed from 43 road-crash attributes, using non-standardized data with natural values to produce a comprehensive cluster model with human decipherable reports. The model was generated by the EM algorithm as a testing set, using 5,695 randomly selected instances of the total 16,750 crashes. The model develops cluster distributions of crashes for the whole main roads network and nicely resolved cluster groups with differing crash profiles. Results of the case study are shown in the appendix (Table 10.2).

The geographic distribution of all clusters plotted on latitude and longitude is shown in Fig. 10.1. Brisbane to Cairns defines the eastern seaboard freeways and Mackay/Townsville to Mt Isa shows the central inland road systems. The emergent patterns differentiate rural highways from the more complex city and regional centers.

Tabulation and correlation of the common characteristics of each cluster lead to grouping of common clusters (sample groups in Table 10.1). Analysis showed that clusters were strongly influenced by the road seal type, speed limit, design features such as the number of lanes, single or dual carriageway, as well as the roadway features such as intersections, roundabouts, and so on. Each cluster group maintained a characteristic road segment *average* crash rate ranging from high (8 crashes per year) to low (less than 1 crash per year). Table 10.1 presents two of the five cluster groups devised to demonstrate the interaction between crash types, crash rates, and road attributes.

Cluster distributions were closely related to the distribution of the seal types, roadway features, roles of the roads, and settlement patterns.

The chart showing relative cluster crash characteristics (Fig. 10.2), clearly demonstrates that each road-crash cluster has its own characteristic set of crashes, and our development of individual cluster crash profiles was based on these results.

This outcome demonstrates that clustering method meets the objective of this paper: to develop a crash profile for each of the cluster groups for use in decision support. The following examples illustrate some differences in the clusters.

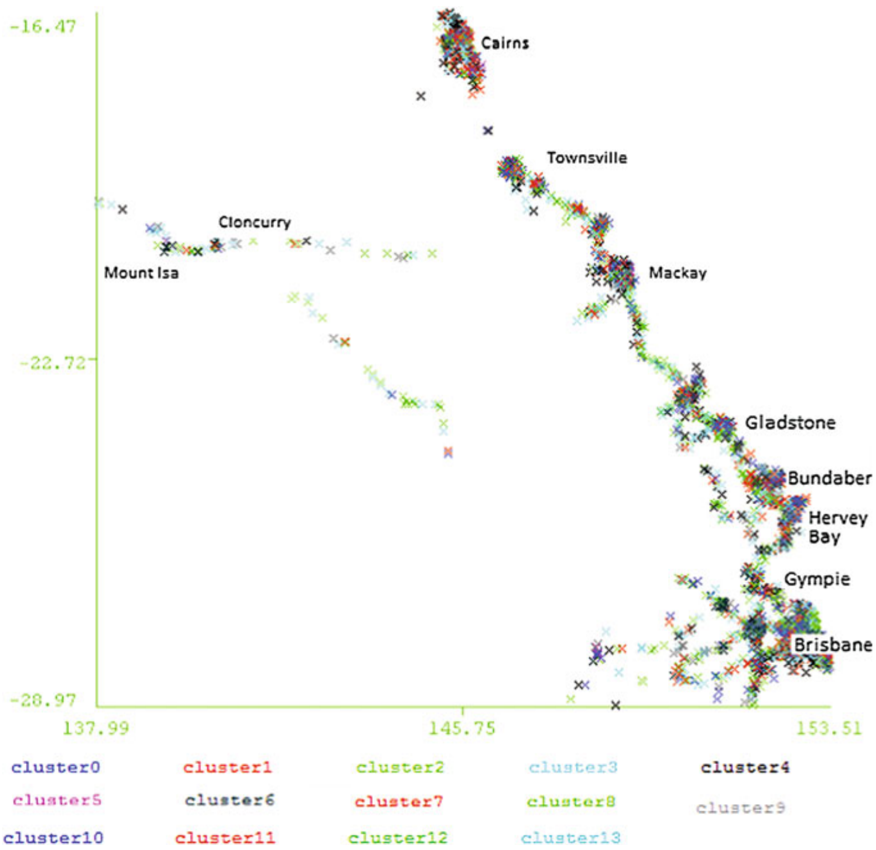
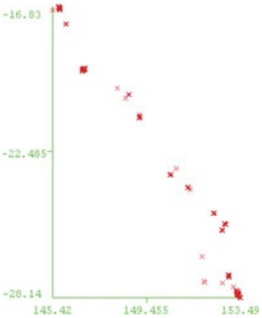
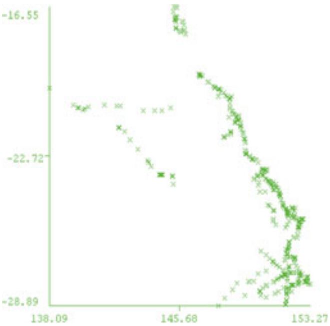


Fig. 10.1 Geographic distribution of main road crash clusters

Freeway and main roads clusters, *cluster* 2 *and* 13, have high levels of head-on crashes as well as having the highest level of vehicles having overturned and crashes involving animals. These road segments, while having relatively low 1 km road segment crash counts, have the most severe results, with *cluster* 13 having all of its crashes involving hospitalization or fatality. *Clusters* 1, 5, *and* 6 have the greatest relative frequency of angle crashes, where road crossings were involved and are among road segment with the highest crash counts.

Individual cluster profiles were used to describe and explain the effective crash characteristics of a roadway; in this study comparing three Brisbane motorways (Fig. 10.3). *Clusters* 0, 2, *and* 12 are predominant clusters of the Ipswich, Gateway and Pacific motorways, and their respective crash types profiles help explain their crash behavior. The seal type at the crash locations for Ipswich and Pacific motorways is predominantly dense graded asphalt, whereas the Pacific highway crash locations are predominantly cement concrete along with other seals. The highway crash profiles appear to have a strong relationship with seal type.

Table 10.1 Samples cluster groups devised from the fourteen clusters in the model

Sample cluster geographic distribution	I km road segment cluster profiles
	Cluster group A: Main road/highway, proximity to city. Clusters 1, 5, 6 Road profile: Seal type of dense graded asphalt, with daily traffic rates of between 13k and 17k vehicles per day. Predominantly single carriageway, high percentage of divided road of both 2 and 4 lane @ 60–80 km/h. Annual average crash counts: are the highest ranging between 5 and 8 crashes per road segment with a maximum of 35. Serious crash averaged: average between 17 and 22 % Crash Profiles: high levels of crossings and t-junctions involved in crashes, with angle and rear collisions being the predominant crash types. Wet crash: low with av. values 3–12 %. Wet crash was a low percentage between 0 and 9 %
	Cluster group D :Freeways and main roads Clusters 2, 13 Road profile: Seal type of spray seal, with low daily traffic rates of between 3.2k and 3.5k vehicles per day. Predominantly single carriageway, very high percentage of divided road and exclusively 2 lane @ 80–110 km/h Annual average crash counts: are the lowest ranging between 0.5 and 0.8 crashes per road segment with a maximum of 32. Serious crash rate is the highest with cluster averages between 33 and 100 % Crash Profiles: highest percentage of road features not involved in crash (97 %) with relative high levels of hit stationary object, overturned, head-on being the predominant crash types.

The Ipswich and Gateway Motorways of dense graded asphalt have predominantly crash *cluster* 0 *and* 12, whereas the Pacific Motorway is mostly *clusters* 3 *and* 12.

These clusters link the road and its attributes with crash characteristics. The clusters and their relationships with the motorways are described (Figs. 10.4, 10.5, 10.6, 10.7).

Cluster 0, a motorway/highway type, found on the Ipswich and Gateway motorways, has the following characteristics; seal of dense graded asphalt, with 92 % being undivided two lane, dual carriageway with speed limit of 100 km/h and average traffic rate of 20k vehicles per day. Count of individual crash instances decisions made by on-site crash assessors indicate that 80 % of crashes are not related to roadway features, with the remainder associated with T junctions, roundabouts, and crossings. Road segment crash count is relatively high, averaging at 4 crashes per year of which a quarter being serious. The crash profile (Fig. 10.4) of this cluster shows that rear-end crashes are the most common by a magnitude of 2, followed by hit fixed obstruction, sideswipe, angle, and sideswipe. The cluster

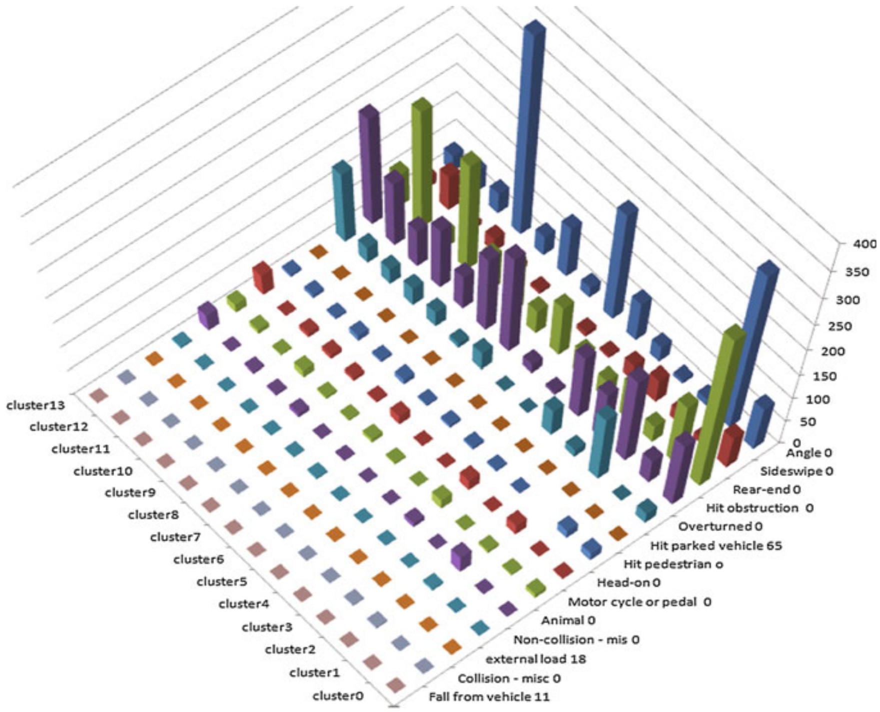


Fig. 10.2 Crash profile of each cluster

profile could be used in decision making to inform signage to alert drivers to the predominance of rear end crashes.

Cluster 12 (Fig. 10.5), a component of all three motorways, shares all attributes in common with cluster 0 (Fig. 10.4), except that the seal type is spray seal. Almost all crashes are assessed as not being related to roadway features. Once again we could apply the crash characteristics provided by the cluster to inform motorists of the impending risks by using appropriate signage.

Cluster 3 (Fig. 10.6), which includes a high proportion of the Pacific Motorway, has a seal of predominantly cement concrete with sections of dense graded and open graded asphalt. The road is 99 % undivided, dual carriageway with over half being 4 lane with the speed limit of 100–110 km/h and average traffic rate of 20 k vehicles per day. Road features, however, are not associated with 95 % of crashes. While the traffic rate of 45k vehicles per day is double that of *cluster* 0, the road segment crash count is substantially lower at 2 crashes per year, of which a quarter are serious. The cluster crash profile (Fig. 10.6) shows that *hit fixed obstruction*, followed closely by *rear-end*, then by *sideswipe*. Similar to cases above, the crash profile may be of help in the development of signage messages to inform motorists.

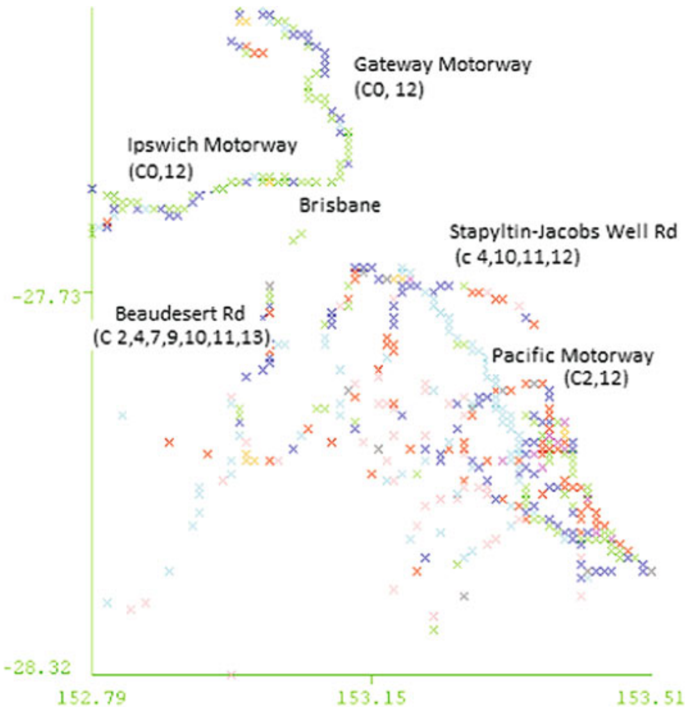


Fig. 10.3 Cluster components of Brisbane motorways and main roads

No.	Label	Count
1	Angle	89
2	Sideswipe	66
3	Rear-end	311
4	Hit fixed Obstruct...	134
5	Overturned	22
6	Hit Parked Vehicle	8
7	Hit Pedestrian	17

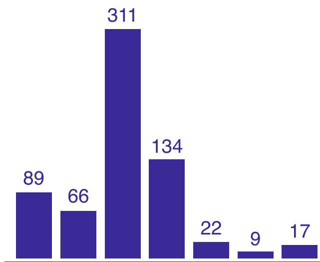


Fig. 10.4 Crash profile of cluster 0 for dense graded asphalt motorway

Main roads, which exhibit a more complex set of conditions, present a more diverse set of crashes than motorways. Figure 10.3 shows that the sections making up the Brisbane to Beaudesert Road include clusters 2, 4, 7, 9, 10, 11, and cluster 13. Motorists are confronted with more frequently changing conditions and risks than with motorway driving. While the road has lower traffic volume at 3.5k vehicles per day, some road segments are prone to a high proportion of serious crashes. The presence of Cluster 13 (Fig. 10.7) is of significance, with all crashes

No.	Label	Count
1	Angle	40
2	Sideswipe	64
3	Rear-end	215
4	Hit fixed Obstruct...	118
5	Overturned	30
6	Hit Parked Vehicle	12
7	Hit Pedestrian	7

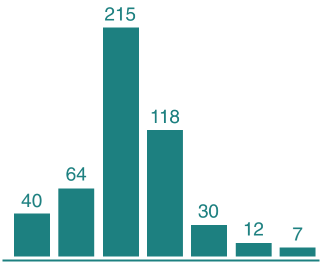


Fig. 10.5 Crash profile of cluster 12 for spray seal motorway

No.	Label	Count
1	Angle	7
2	Sideswipe	46
3	Rear-end	74
4	Hit fixed Obstruct...	96
5	Overturned	15
6	Hit Parked Vehicle	9
7	Hit Pedestrian	2

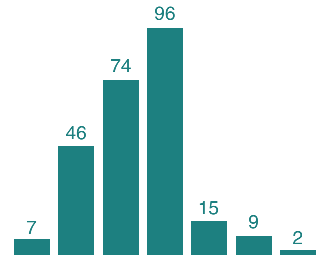


Fig. 10.6 Crash profile for cluster 3

No.	Label	Count
1	Angle	30
2	Sideswipe	19
3	Rear-end	60
4	Hit fixed Obstruct...	197
5	Overturned	130
6	Hit Parked Vehicle	4
7	Hit Pedestrian	5
8	Hit animal incl. rid...	31
9	Head-on	39

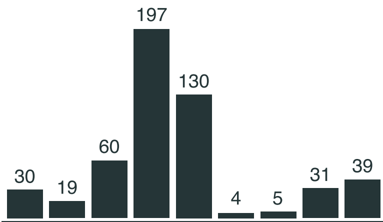


Fig. 10.7 Crash profile for cluster 13

of the cluster being serious, and including a relatively high percentage of potentially fatal crashes, such as *vehicle overturned* and *head-on crash*.

Each of the clusters represent a microcosm that the motorist will encounter during their journey. With information from the clusters, action can be taken based on characteristics of each microcosm to provide appropriate modifications to driver instructions relating to the changing traffic and risk; actions which have the potential to lead to a safer journey for the motorist and potentially contribute to an improvement in road safety.

Incidentally, since the data survey, substantial road works have been conducted to modernize this road section as part of QDTMR's normal development of state roads.

In addition to providing views of road sets, clustering provide new data view for statistical analysis. The clustering algorithm annotates each instance in the dataset with the cluster number, thus giving the analyst groups of homogeneous roads selected from across the whole network. These new groups of instances can be analyzed using statistical tools in the knowledge extraction process [4].

10.5 Conclusion

The paper provides an applied data mining methodology to develop a system of benchmarked clusters for a whole road network, and demonstrates the value of the outcome with a case study. This paper demonstrated the benefit of applying clustering within the road management and road safety domains by developing and applying cluster-based crash profiles to describe the crash characteristics of road sections associated by their cluster. This information can be applied in new ways to make decisions about reducing road crashes.

Resulting clusters were differentiated by value ranges within the road attributes, the crash types and the annual crash rates, and this information was used to develop crash profile for each group of roads. The study demonstrates that the crash profiles provide meaningful information, useful in decision making, shown in an example by identifying and applying the predominant crash type for a section of homogeneous road segments selected from the whole road network. Information including risk and driving strategies could be generated and be presented to the motorist using roadway signage. These driver messages could be made specific to time of day, time of year and the prevailing weather conditions.

Further, this study describes significant differentiation of attribute values between clusters, allowing analysis of the road characteristics and the individual crash-road instance within and between cluster allocations. In future work this differentiation could be used to support higher order knowledge development. Standard statistical methods can be used in analysis to support hypothesis development, which in turn could be used to support road design and maintenance procedures focused on road safety.

Acknowledgments This paper was developed within the CRC for Infrastructure and Engineering Asset Management (CIEAM), established and supported under the Australian Government's Cooperative Research Centres Programme. The authors gratefully acknowledge the financial support provided by CIEAM. This CIEAM study is a component of a project involving QUT and QDTMR in the development of a skid resistance decision support system for road asset management [11, 13–15]. Queensland Department of Transport and Main Roads provided the road and crash datasets. Data mining operations and presentations were performed in the WEKA and SAS platforms, and substantial computer processing was performed on the QUT High Performance Computing facility (HPC). The views presented in this paper are of the authors and not necessarily the views of the organizations.

10.6 Appendix 1

Table 10.2 Case study cluster details

A: Proximity to City						Geometry		R.way feature		Crash type						Roadway Design				Interest	
Cluster	Segment Total	4year Crash Count mean	AADT mean	Predominant Seal	Predominant Road type	%Curved	%Rolling/ Mountain	%Not Applicable	%Crossing Tjunction	%Angle	%Side Swipe	%Rear End	%Hit Obstruction	%Head On	%Bike m. cycle	%Dual Carriageway	%Divided Road	%2 lane	%4 lane	%Serious	%Wet
1	523	19.0	13,114	DGA	MAIN/HIGHWAY	9.4	6.5	8.8	76.3	59.1	2.3	24.3	9.4	0.2	0.8	85.5	13.6	100	0	22	3
5	140	18.3	12,994	DGA	MAIN/HIGHWAY	7.9	7.1	14.3	71.4	53.6	3.6	34.3	5.7	0.0	0.0	69.3	23.6	80	1	17	8
6	343	32.8	15,168	DGA	HIGHWAY/Main	5.5	5.5	17.8	77.0	59.8	2.6	28.3	5.2	0.3	1.2	25.1	53.6	42	44	16	0
B:	Urban & Regional																				
10	774	8.6	7,947	SS	HIGHWAY /Main	16.4	13.7	31.5	60.5	45.6	3.2	25.7	14.6	1.2	1.8	17.3	60.7	100	0	28	4
11	202	7.1	6,729	SS	HIGHWAY /Main	37.1	29.2	70.8	26.2	18.8	4.5	14.9	34.7	4.5	1.5	23.8	63.9	100	0	26	18
4	337	3.9	3,516	SS	H.WAY/Main/Sec	57.9	43.3	86.4	12.5	8.9	8.9	11.0	37.1	6.5	5.9	1.2	9.8	96	0	35	11
7	337	6.0	5,398	SS	HIGHWAY/Main	57.3	35.0	81.3	16.6	5.9	3.0	13.6	55.5	5.3	2.4	9.2	81.9	100	0	23	100
9	244	10.5	8,370	SS	HIGHWAY /Main	23.4	13.5	84.0	13.1	14.8	4.5	27.5	26.2	2.9	2.0	12.3	79.1	82	8	27	10
C:	City Motorway/Highway					0.0															
3	256	9.3	44,598	DGA	H.WAY/Urban Art	10.5	21.1	94.5	4.3	2.7	18.0	28.9	37.5	0.4	0.8	100.0	0.4	18	52	26	18
0	663	14.5	20,080	DGA	HIGHWAY/Main	15.7	12.5	78.0	15.7	13.4	10.0	46.9	18.7	0.5	1.5	92.0	9.8	92	3	23	4
12	504	18.0	20,106	SS	URB Art/Hwy/Main	17.7	14.5	83.7	8.7	7.9	12.7	42.7	23.4	0.2	1.8	94.6	6.9	100	0	18	3
8	375	18.5	17,858	DGA	HWY/MAIN/UrbArt	23.2	14.7	50.7	38.7	25.3	1.6	26.4	37.6	1.1	0.8	74.9	20.3	85	9	22	100
D:		Regional Highway/Main				0.0															
13	544	2.9	3,555	SS	HIGHWAY/Main	16.4	13.4	96.1	2.2	5.5	3.5	11.0	0.4	7.2	2.9	0.7	95.8	100	0	100	0
2	453	2.4	3,216	SS	HIGHWAY/Main	34.0	22.7	97.6	1.8	2.9	5.3	7.5	38.2	4.6	1.3	4.0	90.7	100	0	33	13

References

1. World Health Organization (2002) Global health: today's challenges. The World Health Organization. Available from: http://www.who.int/whr/2003/en/Chapter1.pdf. Cited 11 Feb 2011
2. Queensland-Fire-and-Rescue-Service (2002) Firefighters called to record number of road crashes. Queensland Fire and Rescue Service, Queensland Government (Brisbane) [serial online]. 18 Nov 2002. Available from: http://www.fire.qld.gov.au/news/view.asp?id=207. Cited 2008 October
3. Queensland Department of Transport Main Roads (2011) Road planning and design manual, design philosophy. Queensland Government, Transport and Main Roads (Brisbane) [serial online]. 11 Feb 2011. Available from http://www.tmr.qld.gov.au/Business-and-industry/Technical-standards-and-publications/Road-planning-and-design-manual.aspx. Cited 02 Aug 2010
4. Huang H, Chin H, Haque M (2009) Empirical evaluation of alternative approaches in identifying crash hot spots. Trans Res Rec: J Trans Res Board 32–41
5. ARRB Group (2011) Tools for improving road safety. ARRB Group. Available from: http://www.arrb.com.au/Safe-Systems/Tools-for-improving-road-safety.aspx. Cited 1 Sept 2001
6. Depaire B, Wets G, Vanhoof K (2008) Traffic accident segmentation by means of latent class clustering. Accid Anal Prev 40(4):1257–1266
7. Alzghoul A, Löfstrand M (2011) Increasing availability of industrial systems through data stream mining. Comput Ind Eng 60(2):195–205
8. Pande A, Abdel-Aty M (2009) Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. Saf Sci 47(1):145–154
9. Zhao Z, Jin X, Cao Y, Wang J (2010) Data mining application on crash simulation data of occupant restraint system. Expert Syst Appl 37(8):5788–5794

10. Anderson TK (2009) Kernel density estimation and K-means clustering to profile road accident hotspots. Accid Anal Prev 41(3):359–364
11. Emerson D, Nayak R, Weligamage J (2011) Using data mining to predict road crash count with a focus on skid resistance values. In: Proceedings of the 3rd international surface friction conference. Gold Coast, p 18
12. Qin X, Ivan JN, Ravishanker N (2004) Selecting exposure measures in crash rate prediction for two-lane highway segments. Accid Anal Prev 36(2):183–191
13. Nayak R, Emerson D, Weligamage J, Piyatrapoomi (2011) Road crash proneness prediction using data mining. In: Proceedings of the extending database technology. Uppsala, Sweden. p 521–526
14. Nayak R, Emerson D, Weligamage J, Piyatrapoomi N (2010) Using data mining on road asset management data in analysing road crashes. Proceedings of the 16th Annual TMR Engineering and Technology Forum, Brisbane
15. Emerson D, Nayak R, Weligamage J, Piyatrapoomi N (2010) Identifying differences in wet and dry road crashes using data mining. Proceedings of the Fifth World Congress on Engineering Asset Management, Brisbane

Chapter 11
M-ary Trees for Combinatorial Asset Management Decision Problems

S. Chakraborty, C. Fidge, L. Ma and Y. Sun

Abstract A novel *m*-ary tree-based approach is presented to solve asset management decisions which are combinatorial in nature. The approach introduces a new dynamic constraint-based control mechanism which is capable of excluding infeasible solutions from the solution space. The approach also provides a solution to the challenges with ordering of assets decisions.

11.1 Introduction

Asset management is about finding the right balance between cost, performance, and risks to any asset. Asset management helps us to align corporate goals with maintenance spending and to draw up future asset plans [1]. In large organizations with a large number of assets, the task of the asset manager is a challenging one. In this paper, we focus on a combinatorial decision support problem relevant to asset maintenance. The problem concerns a group of assets, each with multiple maintenance options, for each of which the decision maker needs to select one option. While making these option selections, the decision maker must find the combination

S. Chakraborty (✉) · C. Fidge
Computer Science, Science and Technology, Queensland University of Technology, Queensland QLD 4101, Australia
e-mail: subrata.chakraborty@qut.edu.au

C. Fidge
e-mail: c.fidge@qut.edu.au

L. Ma · Y. Sun
Engineering Systems, Faculty of Built Environment and Engineering, Queensland University of Technology, Queensland QLD 4101, Australia
e-mail: l.ma@qut.edu.au

Y. Sun
e-mail: y3.sun@qut.edu.au

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_11,

that satisfies all technical and business constraints best. The most obvious way to handle this problem is to generate all candidate solutions (combinations of options) and then compare them with respect to the decision constraints. In practice, however, this simple, brute-force algorithm is too computationally intensive for large numbers of assets or maintenance options. The number of candidate solutions and required comparisons increases exponentially with the number of assets. In order to solve such an intractable computational problem, we need to significantly reduce the number of potential solutions to be compared by limiting the generation of infeasible solutions.

One of the most popular tree-based approaches in management and decision analysis is the decision tree [2, 3]. Decision trees have been successfully used in asset management [4, 5]. The decision-tree approach is very efficient for decision problems with small numbers of assets. However decision trees are designed for sequentially guiding choices in order to identify a single solution, not for generating large numbers of equally—acceptable solutions.

In the area of financial optimization and asset-liability assessment, stochastic programming models are also popular. The stochastic programming models generally require developing a scenario tree [6–10]. The stochastic scenario tree can be considered as similar to the *m*-ary tree [11].

Binary trees [12] and their variants have been an integral part of computing since the early days. They are widely used in algorithms, programming, data structures, searching, and sorting [13–16]. Sorted binary trees have a natural ability to work efficiently with a large number of variables (decisions) due to their ability to rapidly partition the search space. This capacity makes the binary tree-based approach suitable to deal with combinatorial decision problems. With a binary tree, each parent node has two branches. In the case of asset management decision problems, a parent node can be considered to denote an asset and each of its branches as alternative maintenance options for that asset. In practice, however, an asset may have more than just two maintenance options. To handle this we need to use an *m*-ary tree [11], which is a generalization of binary trees where a parent node has *m* branches.

In the following sections, we first explain the decision problem we are concerned with in mathematical terms followed by the details of an *m*-ary tree suitable for solving the problem. We then present a novel *m*-ary tree-based approach to solve the combinatorial asset management decision problem followed by an industry-based case study to demonstrate the applicability of the new approach.

11.2 Problem Statement

The assumption is that each decision must contain one and only one maintenance option for each asset. Given a set of assets N, each with a set of an $(n = 1, 2 \ldots, N)$ maintenance options, the number of possible decisions D can be expressed as the product:

$$D = \Pi_{n=1}^{N} a_n. \tag{11.1}$$

The assumption is that the maintenance options for any asset are independent of the maintenance options of other assets, i.e., the choice of one maintenance option for an asset does not constrain the options for another asset. The number of possible decisions D includes all potential solutions including infeasible ones, i.e., options that are invalid because they individually violate some overall constraint or because they are incompatible with other options. With a large number of assets and maintenance actions, the value of D can be astronomical. For example, consider 10 assets with 4 maintenance actions for each and the number of possible maintenance decisions will be 410 (=1,048,576). Generally, large organizations make maintenance decisions for hundreds or even thousands of assets. The number of possible maintenance decisions is extremely large by nature and it is a classic combinatorial problem. In order to find the best possible decision among all the possible decisions, we can use a simple brute-force technique as follows:

Step 1: Create all possible maintenance decisions.
Step 2: Compare the candidate solutions based on some constraint like cost, time, importance, etc.
Step 3: Exclude non-feasible solutions and choose one of the remaining acceptable solutions.

In theory, the above technique seems easily achievable. But it is impractical to use it to solve real life problems due to its very high computational requirements. Furthermore, in practice, decision makers are not always after the best solution, rather they are satisfied with any solution that meets their criteria. Limited amounts of optimization testing are usually conducted to check the validity of the chosen decision. The algorithm presented in this paper provides a solution to this combinatorial decision problem, which can:

- Reduce the number of maintenance decisions generated.
- Solve the issue of maintenance decision conflicts dynamically.
- Apply various decision constraints effectively.

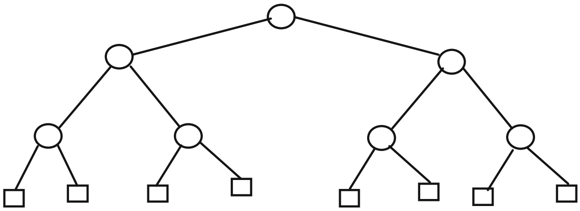
11.3 *M*-ary Trees

A binary tree as shown in Fig. 11.1 contains parent nodes (nodes with successors) and child nodes (nodes without any successor). Binary trees are a rooted tree and recursive in nature. In a fully balanced binary tree [14] with tree height h, the number of leaves L can be calculated as

$$L = 2^h. \tag{11.2}$$

Binary trees can be generalized as m-ary rooted trees where $M \geq 2$ is a fixed integer [11]. The number of leaves in an m-ary tree can be obtained by modifying Eq. (11.2) as

Fig. 11.1 Binary tree



$$L = m^h. \tag{11.3}$$

Equation (11.3) can be generalized as

$$L = \Pi_{i=1}^{h} m_i. \tag{11.4}$$

In an m-ary tree, all the parent nodes in each level of the tree have a fixed number of child nodes m. But Eq. (11.4) is capable of calculating the number of leaves in a tree even if the number of child nodes is different for the parent nodes of any particular level of the tree (m may be different for different levels of the tree).

Let us consider an asset management scenario for Eq. (11.4). The set of assets N represents the height of the tree h and the set of maintenance options $a_n(n = 1, 2..., N)$ represents the number of child nodes m_i at each level of the tree.

The number of decisions D, which consists of one maintenance option for each asset can be compared with the number of leaves L. Based on this analogy we can establish that Eq. (11.1) represents the m-ary tree where m is variable across assets.

11.4 Our Approach

Step 1: Get asset list
Step 2: Get maintenance options for each asset
Step 3: Get attributes for each maintenance option
Step 4: Get filtering criteria
Step 5: Get decision constraints
Step 6: Add default maintenance option
Step 7: Filter assets
Step 8: Control solution space
Step 9: Compare decisions.

The details of the approach are discussed as follows:

Step 1–5: Inputs

The approach requires the following inputs:

- An asset inventory listing all assets A
- A set of maintenance options for each asset $a_n(n = 1, 2..., A)$

- Attributes of each maintenance option that influence the decision outcome. Examples of attributes can be time, cost, safety, operational significance, etc.
- A set of asset filtering criteria F. Examples of filtering criteria can be geographical location, known operational constraints, time of the year, etc.
- Decision constraints C such as
  - Incompatible maintenance options.
  - Limit constraints (time, cost, resources, safety, etc.).

Step 6: Add Default Maintenance Option

In this stage, a default maintenance option is added to each asset. The default option is considered as 'No Action' for the asset. It has no cost associated with it and it does not require any time or any other resources. The key benefit is that it allows reaching all possible solutions without creating multiple tree structures for making asset decisions in different order.

Considering Fig. 11.2, which shows the potential search space generated when sequentially making decisions for an inventory containing three assets. If we want to make a decision consisting of options from assets A2 and A3, we must make a decision for asset A1 as well. The only way to make any decision with options from A2 and A3 only and not with A1 is to restructure the tree with starting node being A2 or A3. Thus to reach all possible decision scenarios, we may need to develop multiple tree structures. The addition of the default option to each asset provides access paths to all possible solutions without the need to create several tree structures for each possible ordering of the assets. Figure 11.3 shows the improvement from Fig. 11.2 that is capable of providing access to all possible decisions. X represents the default option. When selected, it indicates that no initially given option will be selected for that asset and yet it will provide access to other asset options to the lower levels of the tree.

Step 7: Filter Assets

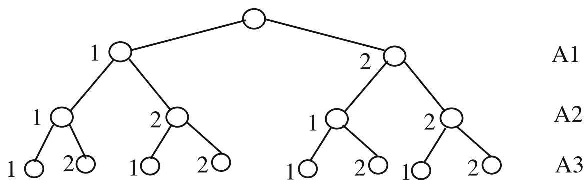
From the asset inventory A, the set of assets N relevant to current decision making can be filtered based on given filtering criteria F as the elements x for which the filtering criteria are true, i.e.,

$$N = \{x | x \in A \wedge F(x)\}. \tag{11.5}$$

Step 8: Control Solution Space

The complete solution space contains all the possible decisions (combinations of maintenance options for each asset) D, i.e., all the paths from the root of the state space tree to its leaves. The control stage is a directed search through the

Fig. 11.2 Three assets with two options each



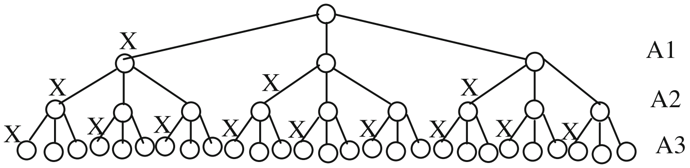


Fig. 11.3 Three assets with two options and added default option

complete solution space D excluding branches consisting of infeasible decisions. This process dynamically reduces the solution space by not generating infeasible decisions and their offspring. The set of feasible decisions d can be obtained as the set of elements x for which the decision criteria hold, i.e.,

$$d = \{x | x \in D \wedge C(x)\} \tag{11.6}$$

Step 9: Compare Decisions

The comparison is essentially a sorting operation for the set of feasible decisions d based on comparison criteria, which may include control criteria thresholds. Due to the dynamic control applied in Step 2, the number of comparisons required can reduce significantly depending on the decision problem.

11.5 Case Analysis

In this section, we present a small example to demonstrate the potential gains provided by our approach.

11.5.1 Case Description

A large Australian power generating corporation currently maintains thousands of assets as part of their regular operation. They often face the challenge of deciding which assets to maintain, when to maintain them, and how several assets can be grouped together for maintenance. The major constraints and issues they must consider include available maintenance time, costs, and remaining time for safe operation, resource availability, and operational criticality.

In order to illustrate the approach, we will solve a much downsized version of the actual decision problem. The solution will be obtained by applying a single decision constraint. The inputs to the algorithm are as follows:

- The asset inventory listing all available assets A, which contains several thousand assets.
- The set of available maintenance options for each asset $a_n (n = 1, 2, \ldots, A)$ whose number varies for different assets. The operational criticality of each asset is also available.

Table 11.1 Assets, maintenance options, and cost

Asset	Maintenance option (O)	Cost ($100 k)
A1	1	0
	2	10
	3	40
A2	1	0
	2	5
	3	15
A3	1	0
	2	18
	3	30
A4	1	0
	2	20
	3	35

- The available attributes of each maintenance option that influence the decision outcome. The most significant attributes considered are cost, safety, and time.
- The set of asset filtering criteria F. Major filtering is done based on operational area segments in the plant and available major downtime.
- Decision constraints C such as:
  - Incompatible maintenance constraints: Not considered currently.
  - Limiting constraints: Cost and time limits are available.

11.5.2 Solution Using the Algorithm

Steps 1–6 were applied to get all the required inputs and to add the default option for each asset. The default option is denoted as option 1 for each asset.

Step 7: Filter Assets

Based on the power plant segmentation we were able to filter the asset inventory. The filtered asset set N contains a few hundred assets. For the sake of this example, let us assume that we were able to subdivide one of the plant segments into a much smaller area consisting of just four assets. Each asset has three maintenance options including the default option added using Step 6. The related cost attributes of each option are presented in Table 11.1.

Step 8: Control

With four assets having three maintenance options each, if we apply the brute-force technique we will be generating 81 decisions as shown in Fig. 11.4.

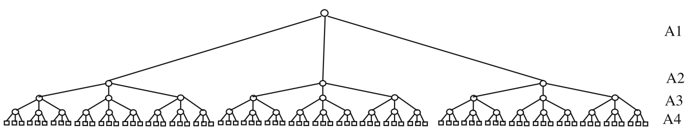


Fig. 11.4 Tree for complete solution space

Table 11.2 Decisions and cost

Decision	Asset options	Cost ($100 k)
D1	A11 A21 A31 A41	0
D2	A11 A21 A32 A41	18
D3	A11 A21 A33 A41	30
D4	A11 A22 A31 A41	5
D5	A11 A22 A32 A41	23
D6	A11 A22 A33 A41	35
D7	A11 A23 A31 A41	15
D8	A11 A23 A32 A41	33
D9	A11 A23 A33 A41	45
D10	A12 A21 A31 A41	10
D11	A12 A21 A32 A41	28
D12	A12 A21 A33 A41	40
D13	A12 A22 A31 A41	15
D14	A12 A22 A32 A41	33
D15	A12 A22 A33 A41	45
D16	A12 A23 A31 A41	25
D17	A12 A23 A32 A41	43
D18	A12 A23 A33 A41	55
D19	A13 A21 A31 A41	40
D20	A13 A21 A32 A41	58
D21	A13 A21 A33 A41	70
D22	A13 A22 A31 A41	45
D23	A13 A22 A32 A41	63
D24	A13 A22 A33 A41	75
D25	A13 A23 A31 A41	55
D26	A13 A23 A32 A41	73
D27	A13 A23 A33 A41	85
D28	A11 A21 A31 A42	20
D29	A11 A21 A32 A42	38
D30	A11 A21 A33 A42	50
D31	A11 A22 A31 A42	25
D32	A11 A22 A32 A42	43
D33	A11 A22 A33 A42	55
D34	A11 A23 A31 A42	35
D35	A11 A23 A32 A42	63
D36	A11 A23 A33 A42	65
D37	A12 A21 A31 A42	30
D38	A12 A21 A32 A42	58
D39	A12 A21 A33 A42	60
D40	A12 A22 A31 A42	35
D41	A12 A22 A32 A42	53
D42	A12 A22 A33 A42	65
D43	A12 A23 A31 A42	45

(continued)

Table 11.2 (continued)

Decision	Asset options	Cost ($100 k)
D44	A12 A23 A32 A42	63
D45	A12 A23 A33 A42	75
D46	A13 A21 A31 A42	60
D47	A13 A21 A32 A42	78
D48	A13 A21 A33 A42	90
D49	A13 A22 A31 A42	65
D50	A13 A22 A32 A42	83
D51	A13 A22 A33 A42	95
D52	A13 A23 A31 A42	65
D53	A13 A23 A32 A42	93
D54	A13 A23 A33 A42	105
D55	A11 A21 A31 A43	35
D56	A11 A21 A32 A43	53
D57	A11 A21 A33 A43	65
D58	A11 A22 A31 A43	40
D59	A11 A22 A32 A43	68
D60	A11 A22 A33 A43	70
D61	A11 A23 A31 A43	50
D62	A11 A23 A32 A43	68
D63	A11 A23 A33 A43	80
D64	A12 A21 A31 A43	45
D65	A12 A21 A32 A43	63
D66	A12 A21 A33 A43	75
D67	A12 A22 A31 A43	50
D68	A12 A22 A32 A43	68
D69	A12 A22 A33 A43	80
D70	A12 A23 A31 A43	70
D71	A12 A23 A32 A43	78
D72	A12 A23 A33 A43	90
D73	A13 A21 A31 A43	75
D74	A13 A21 A32 A43	93
D75	A13 A21 A33 A43	105
D76	A13 A22 A31 A43	80
D77	A13 A22 A32 A43	98
D78	A13 A22 A33 A43	110
D79	A13 A23 A31 A43	90
D80	A13 A23 A32 A43	108
D81	A13 A23 A33 A43	120

The decisions and associated costs are listed in Table 11.2, which is our complete solution space for this decision problem.

To apply our new approach, let us assume we have a budget limit of $3,000,000. Our constraint here is the cost constraint for control purposes. Based on this constraint, if we start generating the tree we will only generate the non-circled region as shown in Fig. 11.5. The decisions that are excluded from this

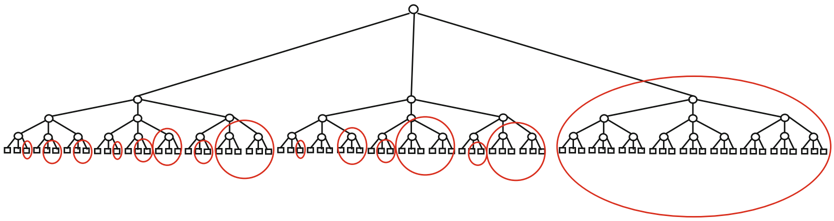


Fig. 11.5 Decision tree using new algorithm

Table 11.3 Feasible decisions and cost

Decision	Asset options	Cost (AUD in 100,000)
D1	A11 A21 A31 A41	0
D2	A11 A21 A32 A41	18
D3	A11 A21 A33 A41	30
D4	A11 A22 A31 A41	5
D5	A11 A22 A32 A41	23
D7	A11 A23 A31 A41	15
D10	A12 A21 A31 A41	10
D11	A12 A21 A32 A41	28
D13	A12 A22 A31 A41	15
D16	A12 A23 A31 A41	25
D28	A11 A21 A31 A42	20
D31	A11 A22 A31 A42	25
D37	A12 A21 A31 A42	30

control phase without even generating them are circled in red. The feasible decisions generated are listed in Table 11.3.

Step 9: Compare Decisions

An appropriate sorting technique can be used to find the best decision among the feasible solutions. Comparing Tables 11.2 and 11.3, we can see that if we use the brute-force technique we will end up comparing 27 decisions to each other. On the other hand with the new algorithm we need to compare only the 10 feasible decisions. This certainly reduces the computational burden.

11.6 Conclusion

The new approach proposed in this paper can be of practical use in the case of asset management decision problems where a large number of assets are involved. The dynamic control does require some additional computations but the benefits far outweigh this extra effort by eliminating large numbers of infeasible decisions without even assessing them. Performance of the algorithm is dependent on the shape of the tree and the proportion of infeasible solutions. Decision problems with higher numbers of infeasible solutions will achieve better efficiency with this algorithm. Further comparative studies are being carried out to understand the actual computational efficiency of the algorithm.

References

1. Brown RE, Spare JH (2004) Asset management, risk, and distribution system planning. Power Systems Conference and Exposition, 2004. IEEE PES 3:1681–1686
2. Quinlan JR (1986) Induction of decision trees. Mach Learn 1:81–106
3. Quinlan JR (1999) Simplifying decision trees. Int J Hum Comput Stud 51(2):497
4. Emerson D, Nayak R, Weligamage J, Piyatrapoomi N (2011). Identifying differences in wet and dry road crashes using data mining. In: J Mathew et al. (eds.) Engineering asset management and infrastructure sustainability: proceedings of the fifth world congress on engineering asset management (WCEAM 2010, Brisbane). Springer, London
5. Sun Y, Ma L, Fidge C (2010) Using decision trees in economizer repair decision making. In: Proceedings of 2010 prognostics and system health management conference. IEEE, Macau, pp 1–6
6. Carino DR, Kent T, Myers DH, Stacey C, Watanabe K, Ziemba WT (1994) Russell-Yasuda Kasai model: an asset-liability model for a Japanese insurance company using multi-stage stochastic programming. Interfaces 24(1):29
7. Consigli G, Dempster MAH (1998) Dynamic stochastic programming for asset-liability management. Ann Oper Res 81:131
8. Kouwenberg R (2001) Scenario generation and stochastic programming models for asset liability management. Eur J Oper Res 134(2):279
9. Kusy M, Ziemba W (1986) A bank asset and liability management model. Oper Res 34(3):356
10. Yu LY, Ji XD, Wang SY (2003) Stochastic programming models in financial optimization: a survey. Adv Model Optim 5(1):1–26
11. Drmota M (2009) Random Trees. Springer, Dordrecht
12. Bayer R, McCreight E (1972) Organization and maintenance of large ordered indexes. Acta Informatica 1(3):173–189
13. Adamson IT (1996) Data structures and algorithms: a first course, 1st edn. Springer, New York
14. Baldwin D, Scragg G (2004) Algorithms and data structures: the science of computing. Charles River Media, Hingham
15. Binstock A, Rex J (1995) Practical algorithms for programmers. Addison-Wesley, Reading
16. Wirth N (1986) Algorithms and data structures. Prentice-Hall, Englewood Cliffs

Chapter 12
Value at Risk Associated with Maintenance of a Repairable System

T. Cheng, M. D. Pandey and J. A. M. van der Weide

Abstract This paper presents the derivation of the probability distribution of maintenance cost of a repairable system modeled as an alternative renewal process and subjected to an age-based replacement policy. The key idea of the paper is to formulate a renewal equation for computing the characteristic function of the maintenance cost incurred in a fixed time interval. Then, the Fourier transform of the characteristic function leads to the complete probability distribution of cost. This approach also enables the derivation of probability distributions of the down time and the number of failures in a given time period. The distribution of the cost can be used to evaluate the value at risk (VaR) and other measures needed for the financial planning of a maintenance program.

12.1 Introduction

The reliability of an engineering system is maintained through various policies, such as age-based replacement [1, 2], block replacement, and condition-based maintenance [3–6]. Optimizing the maintenance cost against the risk of failure is an important topic in the reliability-centered maintenance.

In most of the literature, maintenance optimization is based on minimization of the asymptotic cost rate, because the renewal theorem provides a simple

T. Cheng (✉) · M. D. Pandey
Department of Civil and Environmental Engineering, University of Waterloo, Ontaria, Canada

M. D. Pandey
e-mail: mdpandey@uwaterloo.ca

J. A. M. van der Weide
Department of Electrical Engineering Mathematics and Computer Science, Delft University of Technology, Delft, The Netherland
e-mail: J.A.M.vanderWeide@tudelft.nl

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_12,

expression for its computation. The asymptotic cost rate is equal to the expected cost in one renewal cycle divided by its expected length or duration. However, this simplification may not be realistic for many engineering systems with a relatively short and finite operating life. The expected maintenance cost in a finite time horizon has been discussed in [7–9].

The evaluation of expected cost is reasonable for finding an optimal maintenance policy among a set of alternatives in a relative sense. However, this approach is not informative enough to estimate financial risk measures, such as the value at risk (VaR), which is a *p*th percentile of the distribution of cost defined as

$$VaR_p(C(t)) = \inf\{x : P\{C(t) \le x\} \ge \rho\}. \tag{12.1}$$

For example, 95th percentile means 5 % probability that the maintenance cost would exceed this value.

To evaluate VaR, it is clear that complete probability distribution of maintenance cost is required. The key contribution of this paper is to present the derivation of the probability distribution of maintenance cost by using the concept of characteristic function (CF). The essence of the proposed method is to derive the CF of cost through a renewal equation. Then, the inverse Fourier transform of CF is computed to obtain the probability distribution of cost. To illustrate this method, age-based replacement policy is analyzed.

12.2 Maintenance Model

A component has a random lifetime *L*. The component is repaired upon it fails, called corrective maintenance (CM), or its age exceeds a threshold value of tp, called preventive maintenance (PM). Then the time to repair is given by $X = \min\{L, t_p\}$. The time spent on repair, denoted by *Y*, is also a random variable. Assume that *Y* is independent of L. After repair, the component becomes as good as new. Then the length of one renewal interval, *T*, is equal to $X + Y$. The component service time consists of a series of pairs $\{X, Y\}$ (see Fig. 12.1), also referred to as a stochastic alternating renewal process or ARP [10]. The ARP model has been popularly used to quantify the failure rate and availability of a wide variety of engineering systems.

In this paper, only discrete time is considered. Denote the probability mass functions (PMF) of L and *Y* by $F_L(k)$ and $f_Y(k), k = 1, 2, \ldots,$,respectively. Then the PMFs of *X* and *T* are given by, respectively,

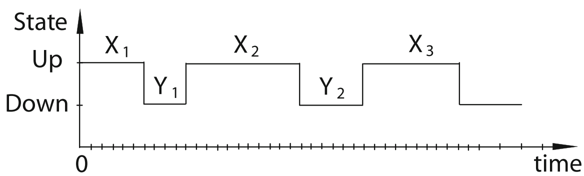


Fig. 12.1 Alternating renewal process model

$$F_y(k) = \begin{cases} f_L(k), & \text{if } k < t_p, \\ \bar{F}_L(t_p - 1), & \text{if } k = t_p, \\ 0, & \text{if } k > t_p, \end{cases} \quad \text{and} \quad f_T(k) = \sum_{x-1}^{k} f_X(x) f_Y(k - x) \tag{12.2}$$

where $\bar{F}_L(k) = 1 - \sum_{j=1}^{k} f_L(j)$ is the survival function (SF) of L. In the following we will always use notations F and $\bar{F}$ with subscript X, Y or T to denote the cumulative distribution function (CDF) and the SF of the associated random variable, respectively, without explicit statements.

In the maintenance model mentioned above, the concerned costs include CM cost c_{CM}, PM cost c_{PM}, and downtime cost. In general $c_{CM} \gg c_{PM}$ due to the loss resulting from component failure. Downtime cost is the economic loss due to unavailability of the component. The cost incurred in the nth renewal interval can be written as

$$C_n = c_D Y_n + \begin{cases} c_{CM}, & \text{if} \quad L_n < t_p, \\ c_{PM}, & \text{if} \quad L_n \geq t_p, \end{cases} \tag{12.3}$$

where c_D is the downtime cost per unit time. Here we assume that if the component fails at time t_p, then a PM cost instead of a CM cost will be incurred. The cost incurred in a finite time interval $(0, t]$ is given by

$$C(t) = \sum_{n=1}^{N(t)} C_n + C\left(S_{N(t)}, t\right), \tag{12.4}$$

where $N(t)$ is the number of complete renewal cycles up to time $S_n = \sum_{k=1}^{n} T_k$ the time to finish the nth renewal cycle, and $C\left(S_{N(t)}, t\right)$ the cost incurred in the time interval from $S_{N(t)}$ to t. Since L and Y are both random variables, $C(t)$is also a random variable. The key objective of this paper is to derive the probability distribution of $C(t)$.

Maintenance cost can be considered as being discrete in a unit of ρ, i.e., $C(t) = 0, \rho, 2\rho, \ldots, n_c\rho$, where $n_c\rho$ is an upper limit of cost such that $\mathrm{P}\{C(t) > n_c\rho\} \approx 0$. Modeling $C(t)$ as a discrete variable is justified on a practical ground. In financial planning, the cost estimates are typically rounded off to 100 or 1,000 of dollars, or any other suitable number depending the cost involved. Therefore, treating the cost as a precise continuous variable is unwarranted. In the above maintenance model, ρ can be taken as the greatest common factor of c_{CM}, c_{PM} and c_D. The upper bound of $C(t)$, $n_c\rho$, can be estimated based on an upper bound number of renewal cycles. The expected number of renewal cycles up to time t is approximately t/μ_T, where

$$\mu_T = \mathrm{E}[T] = \mathrm{E}[X] + \mathrm{E}[Y], \tag{12.5}$$

is the expected length of one renewal interval. The probability that the actual number would exceed $3t/\mu_T$ is expected to be negligible. Therefore, $3\mu_C t/\mu_T$is a conservative upper bound of $C(t)$, where μ_C is the expected cost in one renewal cycle:

$$
\begin{aligned}
h(k) &= \sum_{j=1}^{\min\{\mathrm{k,t_p}\}} [c_D(k-j) + c_R(j)]f_X(j)f_Y(k-j) \\
H(t) &= \sum_{j=1}^{\min\{\mathrm{t,t_P}\}} [c_D(t-j) + c_R(j)]f_X(j)\bar{F}_Y(t-j)
\end{aligned} \tag{12.6}
$$

Thus, n_c can be estimated as

$$n_c \approx \lceil 3\mu_C t/(\rho\mu_T)\rceil, \tag{12.7}$$

where $\lceil * \rceil$ is an integer ceiling function.

12.3 Characteristic Function Method

12.3.1 Basic Idea

Let $\phi(\omega, t)$ be the CF of $C(t)$, i.e.,

$$\phi(\omega, t) = \mathrm{E}\left[e^{iwC(t)}\right], \tag{12.8}$$

where $i = \sqrt{-1}$ is the imaginary number and ω is an argument in the circular frequency domain. Because the cost is a discrete variable, ω will also be a discrete quantity: $0, \Delta\omega, 2\Delta\omega, \ldots, n_c\Delta\omega$. The unit frequency is defined as [11]

$$\Delta\omega = 2\pi/[(n_c + 1)\rho]. \tag{12.9}$$

In general, a frequency is denoted as $\omega_m = m\Delta\omega$. It is recognized that $\phi(\omega, t)$ is the inverse Fourier transform of the PMF of $C(t)$. Then upon $\phi(\omega, t)$is given, the Fourier transform (FT) of $\phi(\omega, t)$will recover the PMF of cost as

$$fc(c, t) = P\{C(t) = c\} = \frac{1}{n_c + 1}\sum_{m=0}^{n_c}\phi(\omega_m, t)e^{-i\omega_m c}. \tag{12.10}$$

The Fast Fourier transform (FFT) algorithm will make the computation of (12.10) very efficient [11].

12.3.2 Characteristic Function of Cost

To derive the CF of cost, the law of total expectation is used to rewrite Eq. (12.8) as

$$\phi(\omega,t) = \sum_{k=1}^{t} \mathrm{E}\left[e^{i\omega C(t)}|T_1 = k\right] f_T(k) + \mathrm{E}\left[e^{i\omega C(t)}|T_1 > t\right] \bar{F}_T(t), \tag{12.11}$$

by conditioning on the length of the first renewal cycle, T_1. The above equation is valid for any $\omega_0, \omega_1, \ldots, \omega_{n_c}$. But the subscript is avoided in this section for sake of brevity and readability of formulas.

When $T_1 = k(\leq t)$ the cost can be written as a sum: $C(t) = C_1 + C(k,t)$, where C_1 is the cost in the first renewal interval and $C(k,t)$ is the cost in the remaining interval. Note that renewal intervals are independent of each other. Then C_1 is conditionally independent of $C(k,t)$. Furthermore, after the first renewal cycle, k becomes a new origin of the subsequent alternating renewal process. Thus, $C(k,t)$ is stochastically the same as $C(t-k)$. Therefore, Eq. (12.11) can be rewritten as

$$\begin{aligned} \mathrm{E}\left[e^{i\omega C(t)}|T_1 = k\right] &= \mathrm{E}\left[e^{i\omega C1} e^{iwC(k,t)}|T_1 = k\right] = \mathrm{E}\left[e^{i\omega C_1}|T_1 = k\right] \mathrm{E}\left[e^{i\omega C(t-k)}\right] \\ &= \mathrm{E}\left[e^{i\omega C_1}|T_1 = k\right] \phi(\omega, t-k). \end{aligned} \tag{12.12}$$

Substituting Eqs. (12.12) into (12.11) leads to the following renewal-type Eq. (12.12)

$$\begin{aligned} \phi(\omega,t) = \sum_{k=1}^{t} &\underbrace{\mathrm{E}\left[e^{i\omega C_1}|T_1 = k\right] f_T(k)}_{f_\phi(\omega,k)} \phi(\omega, t-k) \\ &+ \underbrace{\mathrm{E}\left[e^{i\omega C(t)}|T_1 > t\right] \bar{F}_T(t)}_{G_\phi(\omega,t)}. \end{aligned} \tag{12.13}$$

Functions $f_\phi(\omega,k)$ and $G_\phi(\omega,t)$ are derived in the appendix [see Eqs. (12.20) and (12.21)]. The initial condition of $\phi(\omega,t)$ for any ω is $\phi(\omega,0) = 1$ since $C(0) = 0$. Then $\phi(\omega,1), \phi(\omega,2), \ldots$, can be derived recursively from Eq. (12.13).

Note that $f_\phi(\omega,k)$, $G_\phi(\omega,k)$, and $\phi(\omega,t)$ are all complex variables. Using superscript R and I to denote the real and the imaginary parts, respectively, Eq. (12.13) can be written as

$$\begin{cases} \phi^R(\omega,t) = G_\phi^R(t) + \sum\limits_{k=1}^{t} \left[\phi^R(\omega, t-k) f_\phi^R(\omega,k) - \phi^I(\omega,t-k) f_\phi^I(\omega,k)\right] \\ \phi^I(\omega,t) = G_\phi^I(t) + \sum\limits_{k=1}^{t} \left[\phi^R(\omega, t-k) f_\phi^I(\omega,k) + \phi^I(\omega,t-k) f_\phi^R(\omega,k)\right] \end{cases}. \tag{12.14}$$

Equation (12.14) consists of two coupled recursive equations with initial condition $\phi(\omega,0) = 1$, or $\phi^R(\omega,0) = 1$ and $\phi^I(\omega,0) = 0$.

12.3.3 Expected Cost

Denote $\mathrm{E}[C(t)]$ by $U(t)$. Obviously, $U(t)$ has the initial condition of $U(0) = 0$. Although $U(t)$ can be obtained from $fc(c, t)$ directly, it is more efficient to use the following renewal equation

$$\begin{aligned} U(t) = \sum_{k=1}^{t} U(t-k) f_T(k) + \sum_{k=1}^{t} \underbrace{\mathrm{E}[C_1 | T_1 = k] f_T(k)}_{h(k)} \\ + \underbrace{\mathrm{E}[C(t) | T_1 > t] \bar{F}_T(t)}_{H(t)}, \end{aligned} \tag{12.15}$$

where

$$\begin{aligned} h(k) &= \sum_{j=1}^{\min\{\mathrm{k,t_p}\}} [c_D(k-j) + c_R(j)] f_X(j) f_Y(k-j) \\ H(t) &= \sum_{j=1}^{\min\{\mathrm{t,t_P}\}} [c_D(t-j) + c_R(j)] f_X(j) \bar{F}_Y(t-j) \end{aligned} \tag{12.16}$$

general formula to compute the expected cost in a finite time horizon for any good as new maintenance policies [8, 9].

12.4 Illustrative Example

12.4.1 Data

Both the lifetime L and the repair time Y are Weibull distributed with hazard rate function

$$h(k) = (k/\beta)^{a-1}, \quad (k = 1, 2, \ldots, \beta) \tag{12.17}$$

For the L shape parameter is $\alpha_L = 4$ and the upper bound is $\beta_L = 50$ and for, $Y, \alpha Y = 2$ and $\beta Y = 3$. Then the means and the standard deviations of L and Y are $\mu L = 23.8$, $\sigma L = 6.5$, $\mu Y = 1.9$, and $\sigma Y = 0.7, ,]] >$ respectively. Concerned costs are $c_{CM} = 10, c_{PM} = 3$, and $c_D = 1$. Then the unit cost is $\rho = 1$. The time horizon is $t = 80$ and the replacement age is taken as $t_p = 30$.

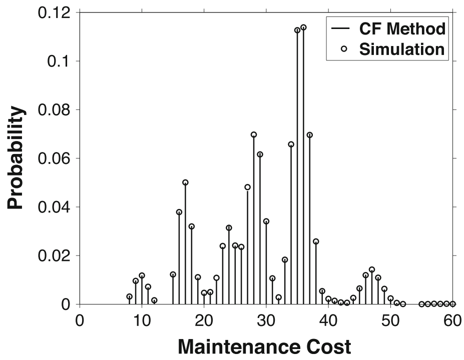


Fig. 12.2 PMF of the maintenance cost $C(t)$

12.4.2 Distribution of Cost

Equations (12.5) and (12.6) lead to $\mu_T = 25.2$ and $\mu_C = 10.5$. Substituting them into Eqs. (12.7) leads to $n_c = 101$. Using Eq. (12.9), the circular frequency, ω, is discretized in steps of $\Delta\omega = 0.06$. The characteristic function of cost, $\phi(\omega, t)$, was computed using Eq. (12.12). Then the PMF of $C(t)$ can be obtained from Eq. (12.10), which is plotted in Fig. 12.2. The distribution is bounded between 0 and 60. The mean and the standard deviation of the maintenance cost are 29.7 and 8.7, respectively.

To verify the results of the proposed method, the Monte-Carlo simulation method was used to obtain the PMF of cost. First, a sample path of the alternating renewal process in time interval $(0, t]$ was generated and the total cost was computed for this sample path. Repeating this over 10^5 simulations, the PMF of $C(t)$ was obtained. The simulated PMF is also presented in Fig. 12.2. It can be seen that the simulated PMF is fairly close to the result obtained from the CF method, which confirms the accuracy of the proposed method. It is interesting that the computational times of these two methods are quite different. The simulation method takes 50 s, while the CF method takes only 1 s on the MATLAB 2010a version.

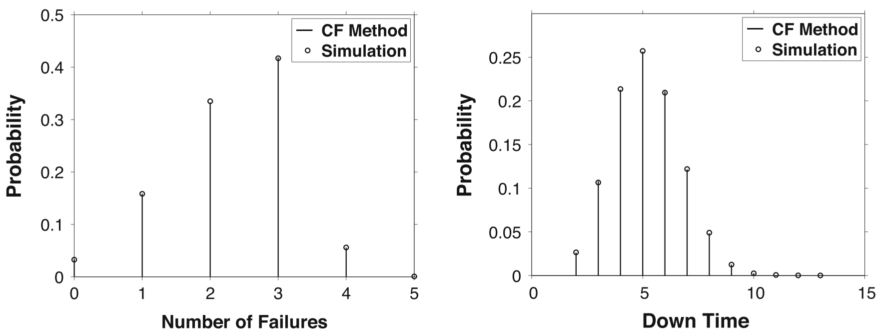


Fig. 12.3 PMFs of number of failures and downtime

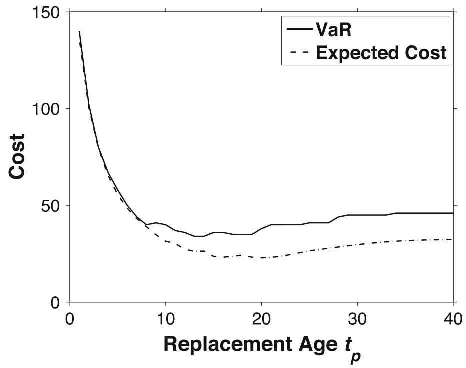


Fig. 12.4 $\mathrm{VaR}_{0.95}$ of cost and the expected cost versus replacement age

12.4.3 Distributions of Number of Failures and Downtime

The maintenance cost formulation is versatile, as it allows to evaluate the PMFs of the number of failures $N_F(t)$, and the downtime in a specified time interval. For example, by setting $c_{CM} = 1$ and $c_{PM} = c_D = 0, C(t)$ becomes synonymous with $N_F(t)$. Similarly, by setting $c_D = 1$ and $c_{CM} = c_{PM} = 0, C(t)$ becomes equivalent to the total downtime up to t. The application of CF method with these cost values will lead to distributions of $N_F(t)$ and downtime, as shown in Fig. 12.3, for $t = 80$ units of time.

12.4.4 Optimization of Replacement Age

The variation of VaR with replacement age is shown in Fig. 12.4. Expected cost is also plotted for comparison. It is seen that minimum VaR is associated with $T_p = 14$. From the point of view of expected cost, it is minimum for $T_p = 20$. Since VaR considers dispersion, it leads to smaller value of T_p.

12.5 Conclusion

This paper presents an efficient method for deriving the probability distribution of maintenance cost in a finite time horizon. The key ideas are to derive the characteristic function (CF) of the maintenance cost via a pair of renewal equations, and then the Fourier transform of CF leads to the probability distribution of the cost.

In the literature, the attention is typically limited to computation of the asymptotic cost rate for optimizing the maintenance policy. The proposed derivation of the complete probability distribution advances the financial risk assessment of a maintenance program. The proposed approach is generic and it can be applied to more involved maintenance policies.

Appendix

Computation of $F_\phi(\Omega, k)$ and $G_\phi(\Omega, t)$

Using the law of total expectation by conditioning on $X_1 f_\phi(\omega, k)$ is obtained as

$$f_\phi(\omega, k) = \sum_{j=1}^{\min\{k, t_p\}} \mathrm{E}\left[e^{i\omega C_1} | X_1 = j, Y_1 = k - j\right] P\{X_1 = j, Y_1 = k - j\} \tag{12.18}$$

Note that the cost incurred in the first renewthat the cost incurred in the first renewal cycle is equal to $c_D Y_1 + c_R(X_1)$, where

$$c_R(X_1) = \begin{cases} c_{CM}, \text{ if } & X_1 < t_p, \\ c_{PM}, \text{ if } & X_1 = t_p, \end{cases} \tag{12.19}$$

is the repair cost. Furthermore, X1 and Y1 are independent of each other. Then Eq. (12.18) gives

$$f_\phi(\omega, k) = \sum_{j=1}^{\min\{k, t_p\}} e^{i\omega[c_D(k-j) + c_R(j)]} f_X(j) f_Y(k - j) \tag{12.20}$$

Similarly, noting that $C(t) = c_D(t - X_1) + c_R(X_1)$ if $X_1 < t < T$ and $C(t) = 0$ if $X_1 > t, G_\phi(\omega, t)$ is obtained as

$$\begin{aligned} G_\phi(\omega, t) = & \sum_{j=1}^{\min\{t, t_p\}} \mathrm{E}\left[e^{i\omega C(t)} | X_1 = j, Y_1 > t - j\right] P\{X_1 = j, Y_1 > t - j\} \\ & + \mathrm{E}\left[e^{i\omega C(t)} | X_1 > t\right] P\{X_1 > t\} \\ = & \sum_{j=1}^{\min\{t, t_p\}} e^{i\omega[c_D(t-j) + c_R(j)]} f_X(j) \bar{F}_Y(t - j) + \bar{F}_X(t) \end{aligned} \tag{12.21}$$

References

1. Christer A (1987) Comments on finite-period applications of age-based replacement models. IMA J Math Manage 1:111–124
2. Nakagawa T (2005) Maintenance theory of reliability. Springer series in reliability engineering. Springer, London
3. van Noortwijk J, van Gelder P (1996) Optimal maintenance decisions for berm breakwaters. Struct Saf 18(4):293–309
4. van Noortwijk J, Klatter H (1999) Optimal inspection decisions for the block mats of the eastern scheldt barrier. Reliab Eng Syst Safety 65:203–211
5. Kallen M, van Noortwijk J (2005) Optimal maintenance decisions under imperfect inspection. Reliab Eng Syst Safety 90(2–3):177–185
6. Nicolai R, van Noortwijk J (2007) A comparison of models for measurable deterioration: An application to coatings on steel structures. Reliab Eng Syst Safety 92(12):1635–1650
7. Christer A, Jack N (1991) An integral equation approach for replacement modeling overfinite time horizons. IMA J Math Appl Bus Ind 3(1):31–44
8. Pandey M, Cheng T, van der Weide J (2010) Finite time maintenance cost analysis of engineering systems affected by stochastic degradation. Proceedings of the Institution of Mechanical Engineers, Part O: J Risk Reliab 225(2):241–250. doi: 10.1177/1748007810393826
9. Cheng T, Pandey M (2011) An accurate analysis of maintenance cost of structures experiencing stochastic degradation. Struct Infrastruct Eng (in press)
10. Birolini A (1999) Reliability engineering: theory and practice. Springer, Berlin
11. Sundararajan D (2001) The discrete Fourier transform: theory, algorithms, and applications. World Scientific, Singapore

Chapter 13
A Proposed Decision-Making Model to Prioritize Building Elements Maintenance Actions Toward Achieving Sustainability in Community Buildings in Australia

P. Kalutara, G. Zhang, S. Setunge, R. Wakefield and H. Mohseni

Abstract Sustainable management of community buildings is a challenging task in Australia. Maintenance and renewal of building assets is a prominent issue as a large number of buildings owned by local councils were built in 1970s and become aged and deteriorated. Each building consists of a massive load of building components which adds complexity into their management. Limited asset management models in favor of buildings' decision making have further widened the gap in finding a reliable decision-making model for building maintenance and renewals. On the other hand, a majority of asset management models available are unable to cope with the uncertainty associated with the data collection which makes the results inconsistent and subjective. This paper presents a useful tool minimizing aforementioned problems and making asset planner's life easier to prioritize maintenance actions. The model is a multi-criteria decision-making model (MCDM) which is combined with two analytical tools, i.e., analytical hierarchical process (AHP) and fuzzy inference system (FIS). The model is based on a four level hierarchical structure, which includes a goal, aspects, criteria, and

P. Kalutara (✉) · G. Zhang · S. Setunge · H. Mohseni
School of Civil, Environmental and Chemical Engineering, RMIT University, Melbourne, Australia
e-mail: pushpitha.kalutara@student.rmit.edu.au

G. Zhang
e-mail: kevin.zhang@rmit.edu.au

S. Setunge
e-mail: sujeeva.setunge@rmit.edu.au

H. Mohseni
e-mail: hessam.mohseni@rmit.edu.au

R. Wakefield
School of Property Construction and Project Management, RMIT University, Melbourne, Australia
e-mail: ron.wakefield@rmit.edu.au

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_13,

attribute factors representing the level one to level four in the hierarchy respectively. Decision Criticality Index (DCI) has been introduced in order to understand the importance of the decision. The concept which is used in the traffic light system is adapted to categorize maintenance options according to color codes depending on DCI value range and the duration of maintenance plan. Example calculations based on case study and hypothetical data have been demonstrated throughout the paper to showcase and validate the model.

13.1 Introduction

Community buildings owned by local councils serve the second largest asset class among all infrastructure assets in Australia based on economic value. That value exceeds more than 15 billion dollars that signifies the importance of having a reliable decision-making model which can address in identifying maintenance actions assuring building elements long life and maintaining their sustainability and the required performances [1]. Reliable and objective understanding of these maintenance actions enable responsible entities to develop appropriate actions and strategies for maintenance, major replacements, refurbishments, and possible future investment on new works [2]. Several software models have been developed in order to address decision making on buildings into their retrofitting aspect [3–5]. The common characteristic of all of them is that the assessment is mainly based on two criteria [6]; (1) residual life which is assessed using deterioration and functional obsolescence of more than 50 different elements, (2) building sustainability which focuses only on specific sustainable areas of the building such as energy consumption, indoor environment, solar heating, etc. Furthermore their assessment process is mainly relied on cost and involved considerable amount of human judgement which causes subjectivity and inconsistency in decisions.

This paper presents a method which aims to address all maintenance actions depending on multi-criteria assessment. According to National Asset Management Strategy group in Australia (NAMS.AU), functionality is a more important concern to align with the needs of the people who use those buildings on a daily basis. Hence functionality is one aspect considered in the proposed decision-making model. Growing interest and current initiatives on improving sustainability in building management has created triple bottom line sustainability factors such as environment, economy and socio as more significant criteria in the decision making on maintenance and renewal work. Therefore, the proposed model evaluates decision making on maintenance in four aspects including functionality, environment, economy, and socio. Moreover, method includes a set of hierarchy where each aspect is divided into several criteria and its attributed factors. Two analytical tools used in multi-criteria decision-making models (MCDM) are combined in the proposed model. They are called analytical hierarchical process (AHP) and fuzzy inference system (FIS); both of them are based on fuzzy logic which enables to minimize subjectivity and inconsistency in decision-making process to a reasonable degree.

The paper is a part of research work which is funded by the Australian Research Council. Six local council organizations in Victoria have made their contribution in the research for establishing a valid sustainable building management model. The outcome of the research enables them to optimize their decision makings with planned maintenance, rehabilitation, and capital expenditure with a realistic understanding of the building deterioration trend. The national benefit includes the significant improvement of service delivery in terms of demand, social aspects, user comfort, risk, sustainability, etc., to community through better design and management of community buildings.

13.2 Literature Review

Previous research works extensively address for decision-making tools on refurbishment of buildings. According to the CRC report of delivering a re-life project [6], two main assessment criteria are identified in the assessment of these tools. Firstly the assessment is taken into consideration of residual life which is determined using its deterioration and functional obsolescence. More than 50 different elements such as windows, roof covering, and façade finish are included in it following degradation and functional obsolescence codes. Second criterion is focused on sustainability limiting into specific concerns including energy consumptions and indoor environment quality. In addition to that for office and hotel buildings, additional modules for solar heating, central air conditioning, pools, etc., are used in the assessment.

Considerable amount of software tools has been developed toward decision making on building retrofits. In most cases radar graph is used to summarize the building deterioration state of the building elements. This is a useful way of visualizing the retrofit cost and identifying the most expensive actions. EPIQR [3], MEDIC [4], TOBUS [5], Office Scorer [7], MAR [8] are some of widely used software tools in relation to decision making on building retrofit. All of these models give an accurate picture about building's existing structural conditions, energy performance, and the final estimate of the total cost. Although these tools are useful valid tools deciding on retrofit of buildings some shortcomings are prevalent inside these tools which are required to address to make them more reliable. Every model is criteria basis in their assessments; though specific criteria are considered in the assessment decisions are eventually made based on solely cost. Also human thinking and judgement are predominant in the assessment process which creates inconsistency and subjectivity on decisions. Exception of MAR model, no other model has made an effort to address other possible maintenance actions on buildings including planned repairs, replacements, etc. On the other hand several other criteria despite of selected criteria used in these models should be considered in the service oriented buildings. Among them functionality plays a key role for acquiring the level of service to customers whom the service is provided to. All of above mentioned models have attempted to focus on sustainability but it has only approached into some specific concerns of environment. Even though partial attention on environment is included in the aforementioned models they have completely ignored the sustainability in socially and economically. Therefore, sustainable triple bottom line factors including environment, economy, and socio should be key concerns on future models.

> Decision making is the study of identifying and choosing alternatives based on the values and preferences of the decision maker. Making a decision implies that there are alternative choices to be considered, and in such a case we want not only to identify as many of these alternatives as possible but to choose the one that best fits with our goals, objectives, desires, values, and so on… [9].

According to Baker et al. [10], decision-making process can be divided into eight steps: (1) Define the problem (2) Determine requirements (3) Establish goals (4) Identify alternatives (5) Define criteria (6) Select a decision-making tool (7) Evaluate alternatives against criteria (8) Validate solutions against problem statement. Fulop [11] has noted the importance of distinguishing problems with single or multiple criteria. He has further explained that if there's a decision problem with a single criterion or a single aggregate measure like cost, then the decision can be made implicitly by determining the alternative with the best value of the single criterion or aggregate measure.

Multi-criteria decision-making models (MCDMs) have been employed for many years as a way to optimize decision outcomes [12, 13]. The basic idea of the approach is to convert subjective assessments of relative importance to a set of overall scores or weights. AHP is one of the more widely applied multi-criteria decision-making methods. AHP has been extensively used by researchers in different applications which are mainly based on pair wise comparison on criteria. According to the Saaty's table there are several number scales ranging from one to nine identifying deferent importance classifications. This has caused some uncertainty and subjectivity in choosing a single number from the scale. To overcome this deficiency Zeng et al. [14] has proposed a table (Table 13.1) which used fuzzy AHP. According to that table, range of values is shown where defuzzified value of them is used as the mean to be used in the AHP calculations.

Table 13.1 Fuzzy AHP pair wise comparison table

Scale	Linguistic definition
1 = (0,1,2)	Equally importance of both elements
3 = (2,3,4)	Moderate importance of one element over another
5 = (4,5,6)	Strong importance of one over another
7 = (6,7,8)	Very strong importance of one element over another
9 = (8,9,10)	Absolute importance of one element over another
2,4,6,8	Intermediate values

Zhang et al. [15] have applied AHP in order to assess the risks in JV projects in China with use of expert's knowledge. According to their method 1st level represents risk condition of joint venture. Next level they identify that risk condition is affected by three major risk groups and each risk group has an impact from several risk factors which use to assess the state of risk group. Likewise level of risky condition has been evaluated using the values of risk groups. Five experts have given their judgement in his case study. Fuzzy multi attribute group decision-making approaches have been used by Kahraman et al. [16] for the decision making of facility location selection. In their study, goal is to select the best catering among the alternative options of using three Turkish firms. Their hierarchy comprise with four levels where goal represents the top level and three catering firms are the bottom level. These levels are combined by two intermediate levels; hygiene, quality of meal, and quality of service have been identified as main parameters influencing to the goal where each parameter is affected by several factors. For example, hygiene has an effect from factors such as hygiene of meal, hygiene of service personnel, and hygiene of service vehicle. In their evaluation process, they have used customers and five experts from Turkish Chamber of Food Engineers. According to their evaluation, Afiyetle has been the best selection of catering firm among selected catering firms. Chiou et al. [17] have adopted AHP for fuzzy multiple criteria decision-making problem to investigate sustainable development of strategies for aquatic product processors in Taiwan. Hierarchy used in this study consists with four main features as goal, aspects, criteria, and finally strategies. Eight strategies have been identified as most beneficial for sustainable development of the fishing industry. Three main aspects are coming under aspects and twelve individual criteria have been formed under criteria. The evaluation method follows first deriving relative weights with respect to each criterion and then calculating corresponding synthetic utility value for each strategy. Later best investment strategies can be decided upon those results. Abdelgawad and Fayek [18] has used AHP in part of the research to measure aggregate index of the impact of failure mode through three influencing parameters of cost, time, and scope. In order to minimize shortcomings using Saaty's table for pair wise comparison, they have adopted the AHP approach proposed by Zeng et al. [14]. Relative importance data are obtained by presenting comparison scale to the senior coordinator and according to his results matrices are formed. From matrix calculations weighting of each parameter is evaluated and multiplying these values with corresponding impact values (obtained as a number ranging from 1 to 10 followed same styles of membership functions for inputs and the output) and total sum generates aggregate impact to the risk from cost, time, and scope.

Probability theory cannot be used to deal with lexical uncertainty because the combination of subjective categories in human decision process does not follow its axioms [19]. Because the fuzzy logic is based on the idea that all things admit of degrees [20] fuzzy inference system (FIS) is a reliable way to deal with decision making which frequently involves human thinking. Furthermore Negnevitsky [20] defines fuzzy inference as a process of mapping from a given input to an output, using the theory of fuzzy sets. FIS is a combination of major four steps such as fuzzification, rule evaluation, aggregation of the rule outputs, and finally defuzzification. Jin et al. [21] have used FIS to model risk allocation in privately financed infrastructure projects. Their fuzzy inference system has six input variables including RM routine, cooperation history, environmental uncertainty, and RM commitment by public partners, RM commitment by private partners and RM mechanism in order to evaluate the output of risk allocation strategy. All input variables and output variable follows Gaussian membership functions varying from 1 to 5 in the variable axis. Membership function of every input variable except RM routine clarified by two linguistic definition terms either low and high or immature and mature. In contrast, membership function of the output variable comprise three linguistic terms in which 1 shows 'retain all' while 3 and 5 represents states of 'equally share' and 'transfer all' respectively. RM routine has also defined in the way of numbering but terms have changed to low, medium, and high. Five experts have been invited to generate fuzzy if–then rules and eliciting the consented opinions of the panel was done by using Delphi procedure. In their model, mat lab software has been used to program fuzzy inference system. Abdelgawad et al. [18] have also used fuzzy inference system in the rest of their study after obtained aggregate impact with respect to cost, time, and scope using AHP process. According to their fuzzy inference system, impact (severity), probability of occurrence and detection/control have considered as three input variables to the output variable of the level of criticality of risk events. Mao [22] has repeatedly applied fuzzy inference system in his study with three level hierarchy systems to estimate values from 1st level to 2nd level and later 2nd level to 3rd level to estimate labor productivity.

13.3 Methodology of the Proposed Model

The methodology starts with a set of hierarchy comprises of four major levels including goal, aspect, criteria, and attributed factors. Attributed factors are used to identify and list out the criteria but do not involve in the model evaluation for decision making. Due to the fact, three major levels as goal, aspect, and criteria are concerned throughout the paper (Fig. 13.1). According to the council interviews and by researching into their practices, four aspects such as functional, environmental, economic, and social were identified as major impacts to decisions in renewals and maintenances. Firstly attributed factors which influence to each aspect were captured through extensive literature review, and their validations were carried out by a pilot survey conducted among six associated councils in the research. Then the factors which have common characteristics and found to be only slight differences among each other were labeled under a criterion. Same method followed for each aspect and, as a result it was able to be condensed into less than five criteria in each aspect. The evaluation process is run in two different methods. The first method focuses on evaluation from criteria level to aspect level whereas the second method targets to evaluate aspect level to goal level.

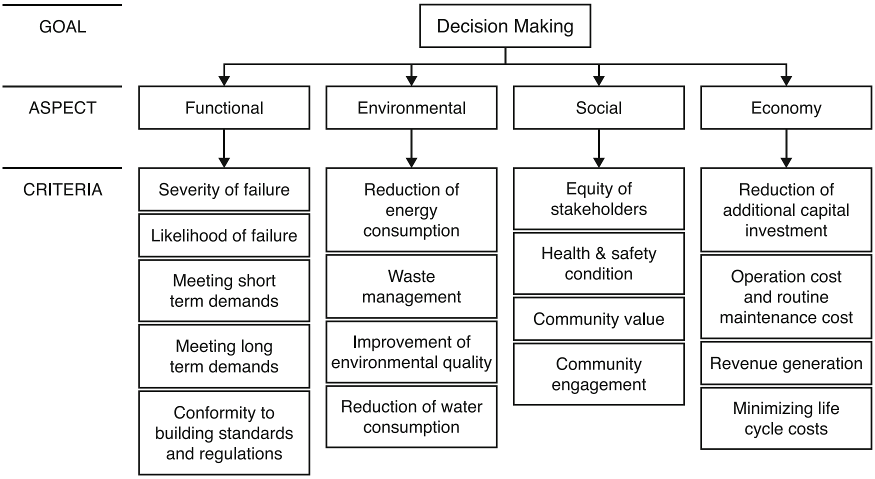


Fig. 13.1 Hierarchy for proposed decision-making model

Fuzzy AHP is employed in the evaluation of the first method. Two important terms were introduced as 'weighting' and 'rating' for each criterion and aspect. Pairs of each criterion in each aspect were compared according to the expert judgements (preferably the building coordinator of the council) which were based on fuzzy AHP relative importance table (Table 13.1). These values form a matrix which enables to calculate weightings of each criterion. The term weighting can be simplified as the importance of the criterion for the aspect. That value is given as a portion of less than one and the sum of all criteria of selected aspect has to be equal to one. Maximum Eigen value is evaluated in order to carry out the consistency index check and, as a result it can be verified the consistency of relative importance data. That index is less than 0.01 is meant that those data are consistent. Following is shown how the weighting and the maximum Eigen value are determined and subsequently check the consistency index. Let P_{ij} denotes relative importance of criterion i over criterion j (This is the mean obtained as the defuzzified value of range of relative importance values). Assume k number of criteria exists, and then Matrix P can be formed according to those values. According to that matrix, Eqs. (13.1), (13.2), and (13.3) can be derived.

Rating is the current level of criterion for a given building element. These data are captured from council's inspection data set. Weighting of a criterion is a fixed value for all elements but the rating of a criterion is changed according to the element. For a given element weighting and rating values can be obtained as explained above. The sum of the multiplication of weighting and rating of each criterion for a given element is the rating of the selected aspect in that element [Eq. (13.4)], where Wi is the weighting and Ri is the rating of criterion i.

$$P = \begin{bmatrix} 1 & P_{12} & . & . & . & P_{1k} \\ 1/P_{12} & 1 & . & . & . & P_{2k} \\ . & . & 1 & . & . & P_{3k} \\ . & . & 1 & . & . & . \\ . & . & . & . & . & . \\ 1/P_{1k} & 1/P_{2k} & . & . & . & 1 \end{bmatrix}$$

Let Wr is the weighting of the criterion r. Then Wr is given by the following equation;

$$W_r = \frac{X}{Y} \quad where\ Y = \sum_{r=1}^{k} X,\ X = \sum_{j=1}^{k} \frac{P_{rj}}{\sum_{i=1}^{k} P_{ij}}. \tag{13.1}$$

$$\lambda_{\max} = \frac{1}{k}[P][W]\left[\frac{1}{W}\right]. \tag{13.2}$$

$$ConsistencyIndex = \frac{\lambda_{\max} - k}{k - 1}. \tag{13.3}$$

$$R_{aspect} = \sum_{i=1}^{k} W_i^* R_i. \tag{13.4}$$

The system structure from aspect level to goal level identifies FIS from the input variables to the output variable. The output variable represents the criticality of the decision making where as the structure comprise of four inputs covering four aspects. All the variables were defined in linguistic terms and formed membership functions accordingly. In the model each variable has five linguistic terms distributed in the range of 1–5 numbering system. Each membership function for linguistic term is considered behaving triangularly. Generation of fuzzy if–then rules is inferred input variables with the output variable. If–then rules are generated with the help of experts. Using all this information a software program was developed using Matlab software. Input values of this program are the values obtained previously via AHP method. Final output for a given element is given accordingly and that value is the decision criticality index of that element. Likewise all the elements are adopted the same procedure and their decision criticality index can be obtained. Once those values are obtained elements can be ranked in the descending order from 5 to 1 where the elements with higher values of DCI are more critical while the elements become less critical as they close to 1 of DCI. Finally those elements can be inserted in several selected options depending on the DCI values and the duration of the maintenance plan.

13.4 An Illustrative Example to Simplify How the Model Runs

Lamps in one of play rooms in a community center of an associated council are considered in the example. The method adopted AHP evaluates its functional aspect according to its criteria. According to Fig. 13.1, five criteria are belonged to the functional aspect. Following figure values were hypothetically taken as relative importance data of their pair wise comparison. F1F2 = 1, F1F3 = 4, F1F4 = 5, F1F5 = 1, F2F3 = 4, F2F4 = 5, F2F5 = 1, F3F4 = 2, F5F3 = 4, and F5F4 = 5. Substituting those values, matrix P as shown in the following can be formed. Then weighting of each criterion can be obtained by the Eq. (13.1). Maximum Eigen value and the consistency index are calculated by using Eqs. (13.2) and (13.3) respectively. All those calculated data are shown in the Table 13.2. The values given by the council for ratings of each criterion for the given element are F1 = 2, F2 = 3, F3 = 1, F4 = 2, F5 = 3. Then overall functional index of the given element can be calculated according to the Eq. (13.4) and the value was taken as 2.50 for the given element.

$$P = \begin{bmatrix} 1 & 1 & 4 & 5 & 1 \\ 1 & 1 & 4 & 5 & 1 \\ 0.25 & 0.25 & 1 & 2 & 0.25 \\ 0.20 & 0.20 & 0.5 & 1 & 0.20 \\ 1 & 1 & 4 & 5 & 1 \end{bmatrix}$$

After each aspect is calculated, method adopted FIS is used to evaluate the decision criticality. In that case, linguistic terms and their membership functions are defined as the first step. All the inputs included functional, environmental, economic, and social aspects are followed the same linguistic terms varying in the same way in the membership functions. Assuming that the council focuses on 10 year maintenance plan, five options were developed in the range of DCI values.

Table 13.2 Obtained criteria results of functional aspects for the given element

	F1	F2	F3	F4	F5
Weighting	0.289	0.289	0.081	0.053	0.289
λ_{max}			5.027		
Consistency index			0.0067		

Table 13.3 Fuzzy if–then rules

IF	Then			
Functionality	Environment	Socio	Economy	Decision criticality
Very good	Very good	Very good	Very good	Very low
Very poor	Very poor	Very poor	Very poor	Very high

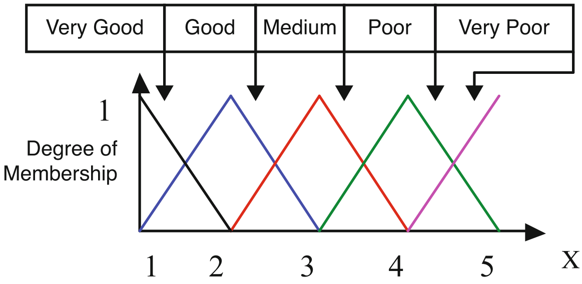


Fig. 13.2 Membership functions of input variables (functional, environmental, social, and economic aspects)

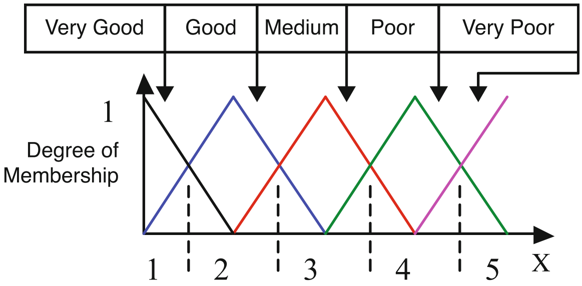


Fig. 13.3 Membership functions of output variable (decision critical)

Table 13.4 Identifying maintenance in different options

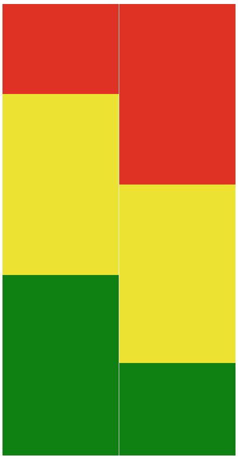
Option (color code)	DCI range	Description
	Between 4.5 and 5.0	Should be attended within 1 year
	Between 3.5 and 4.5	Attending in maintenance between 1 and 3 years
	Between 2.5 and 3.5	Attending in maintenance between 3 and 5 years
	Between 1.5 and 2.5	Attending in maintenance between 5 and 8 years
	Between 1.0 and 1.5	Attending in maintenance between 8 and 10 years

Table 13.5 DCI results, element ranks, and its maintenance options

Building	Floor level	Functional area	Component group	Component type	Component	DCI	Rank	Option
Community center	Ground floor	Play room	Electrical services	Lighting	Lamps	2.50	6	
					Down lights	4.32	2	
					Fittings	4.78	1	
					Exit signs	1.25	8	
					Fluorescent lights	2.64	5	
			External fabrics	Windows and doors	Emergency exit door	3.20	4	
					Automatic opening doors	3.60	3	
					Sliding doors	1.35	7	

Figure 13.2 shows membership functions of input variables and Fig. 13.3 shows the output variable. Table 13.4 shows the options and further describes about them. Fuzzy If–Then rules are generated with the help of experts. Table 13.3 shows two examples for them. Using Mat lab, a program is developed which can be automated the decision criticality index once the input variable values (these are the output values of previous AHP method) are given. Depending on DCI values ranking will be done. Table 13.5 shows DCI values of part of building elements in the associated council and demonstrates the way elements rank further identifying maintenance options accordingly.

13.5 Discussion and Summary

The proposed method follows a new framework capturing important areas in the context of community building management. An extensive effort has put in the research to understand all important parameters which influence the above important areas. The study fills the gap of the fact that lack of attention has paid on building maintenance rather than looking on complete building refurbishment. Furthermore it helps to look on different ways on building maintenance not based on only cost but basis of different criteria. Fuzzy logic has been introduced in two different approaches which make decisions more reliable by minimizing subjectivity and inconsistency in great extent. The main outcome of the model is to derive decision criticalities in building elements which eventually use to rank building elements in the sequence of attending on maintenance. Moreover, the method suggests several maintenance options depending on DCI and the duration of the maintenance plan. Color coded system has introduced to distinguish those options. Using them in building drawings are highly recommended that will be a useful tool to come to fair judgements about building functional areas.

The study is only a solution to the initial stage of decision-making process. Further research is required to understand the final stage of decision making where more attention should be paid on the budget constraints, lifecycle, and remaining useful life which is derived from deterioration curves. Also future study should consider about easy techniques to calculate approximate transition costs in different ratings when the element upgrades. These studies are more related to practical context; therefore undertakings of several case studies are essential in acquiring aforementioned goals.

Acknowledgments The authors would like to express their appreciation to all associated local councils of the research for all their time, effort, and valuable feedback given for the research. We would also like to give our special thanks to the Municipal Association of Victoria for their work to collaborate all other associated councils in group meetings such as progress meetings and workshops. These were highly beneficial to understand local council expectations on decision making and drive the research accordingly. Furthermore Integrate Australia is sincerely remembered for the support provided in the area of computer software development. Finally our appreciation goes to Australian Research Council for the financial support for this research.

References

1. Maunsell (2002) International infrastructure management manual, 2nd edn. Association of Local Government Engineering N. Z. Incorporated and National Asset Management Steering Group, New Zealand
2. Institute of Public Works Engineering Australia—National Asset Management Strategy Committee (2009) Practice note 3: building condition and performance assessment guidelines. IPWEA-NAMS.AU, Australia
3. Flourentzos F, Droutsa K, Wittchen KB (2000) EPIQR software. J Energy Build 31(2):129–136
4. Flourentzos F (2000) MEDIC-a method for predicting residual service life and refurbishment investment budgets. J Energy Build 31(2):167–170
5. Caccavelli D, Gugerli H (2002) TOBUS—a European diagnosis and decision making tool for office building upgrading. J Energy Build 34(2):113–119
6. CRC for construction innovation, web page http://www.construction-innovation.info/index.php?action=article&id=847
7. BRE and dti (2002) Office score: refurbishment support tool for office building. http://www.officescorer.info/
8. Qbuild (2003) Maintenance assessment report for 111 George street. QDPWH, Brisbane
9. Harris R (1998) Introduction to decision making. Available from: http://www.virtualsalt.com/crebook5.htm
10. Baker D, Bridges D, Hunter R, Johnson G, Krupa J, Sorenson K (2002) Guidebook to decision making methods. Department of Energy, USA. Available from: http://emi-web.inel.gov/Nissmg/Guidebook_2002.pdf. Report No.: WSRC-IM-2002-00002
11. Fulop J (2005) Introduction to decision making methods. Laboratory of Operations Research and Decision Systems: Computer and Automation Institute, Hungarian Academy of Sciences 1
12. Saaty TL (1982) Decision making for leaders: the analytical hierarchy process for decisions in a complex world. Lifetime Learning Publishers, Belmont
13. Nagel SS (1984) Public policy: goals, means and methods. St. Martins, New York
14. Zeng J, An M, Smith NJ (2007) Application of a fuzzy based decision making methodology to construction project risk assessment. Int J Project Manage 25(6):589–600
15. Zhang G, Zou XW (2007) Fuzzy analytical hierarchy process risk assessment approach for joint venture construction projects in China. J Constr Eng Manage 133:771–779
16. Kahraman C, Ruan D, Dogan I (2003) Fuzzy group decision—making for facility location selection. Inf Sci 157:135–153
17. Chiou HK, Tzeng GH, Cheng DC (2005) Evaluating sustainable fishing development strategies using fuzzy MCDM approach. OMEGA—Int J Manage Sci 33(3):223–234
18. Abdelgawad M, Fayek AR (2010) Risk management in the construction industry using combined fuzzy FMEA and fuzzy AHP. J Constr Eng Manage 136(9):1028–1036
19. Von Altrock C (1995) Fuzzy logic and neuro-fuzzy applications explained. Prentice Hall, New Jersey
20. Negnevitsky M (2002) Artificial intelligence, a guide to intelligent systems. Addison-Wesley, Essex
21. Jin XH, Doloi H (2009) Modelling risk allocation in privately financed infrastructure projects using fuzzy logic. Comput Aided Civil Infrastruct Eng 24(7):509–524
22. Mao H (1999) Estimating labour productivity using fuzzy set theory. MS Thesis. University of Alberta, Edmonton

Chapter 14
Optimal Design of CSADT with Multiple Stresses

Z. Z. Ge, X. Y. Li and T. M. Jiang

Abstract Most products are affected by multiple stresses simultaneously, so it is necessary to study the optimal method for accelerated degradation testing (ADT) with multiple stresses. This method is proposed for constant stress ADT (CSADT). First, uniform orthogonal test theory is used to determine the combined mode of different stresses. Then stochastic process is used to model the perform degradation of products. Under the constraint of the total experimental cost, the optimum problem is established with the objective that minimizing the asymptotic variance of the estimation of the reliability of the *p*th quantile of product's lifetime under use condition. Optimal test variables are given, including: levels of each stress, total sample size and testing time, and sample size and testing time at each stress combination. Finally, simulation examples are presented to illustrate the proposed method.

14.1 Introduction

Accelerated Degradation Testing (ADT) is proposed as a means to obtain degradation data of products in a short time period and extrapolate the lifetime and reliability of products under use condition.

Boulanger and Escobar [1] did the ADT planning on a class of degradation model. Li [2] proposed a method for planning ADT based on nonlinear mixed effort model. Park et al. [3] discussed the optimal design of an ADT in the case of

Z. Z. Ge · X. Y. Li (✉) · T. M. Jiang
School of Reliability and System Engineering, Beihang University, Beijing 100191, China
e-mail: leexy@buaa.edu.cn

Z. Z. Ge
e-mail: gzz@dse.buaa.edu.cn

T. M. Jiang
e-mail: jtm@buaa.edu.cn

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_14,

destructive measurement. Polavarapu and Okogbaa [4] presented a method for planning ADT by minimizing the interval estimate of MTTF. Yu [5] dealt with the optimal design of an ADT where the degradation rate follows reciprocal Weibull distribution. Tseng et al. [6] proposed an approach of optimal ADT plan when the degradation path follows a gamma process. Wang [7] presented a method for planning accelerated testing based on Monte Carlo simulation. Li and Jiang [8] put forward a method of optimal design for ADT with competing failure model. Ge et al. [9] suggest the optimal ADT plans are devised by minimizing the variance of model parameters.

In these studies, optimal design for ADT was conducted for a case in which only one stress factor was present, but most products are affected by multiple stresses simultaneously. So it is necessary to study the optimal method for ADT with multiple stresses. Pan et al. [10] presented a method for planning ADT with multiple stresses, but in [10] design of ADT was conducted for a case in which stresses were specified without considering combined mode of diffident stresses and levels of these stresses.

In this paper stress will considered as an optimal variable for constant stress ADT (CSADT), with the objective that minimizing the asymptotic variance of the estimation of the reliability of the pth quantile of product's lifetime under use condition and under the constraint that the total experimental cost does not exceed a predetermined budget.

14.2 Testing Profile

For ADT with multiple stresses, we assume that there are L kinds of main stresses, and K stress levels for each stress. S_{l0} is the stress level of normal condition, $S_{l\max}(l = 1,\ldots,L)$ is the stress level of limit conditions, and $S_{lk}(S_{l0} < S_{lk} < S_{l\max})$ $(k = 1,\ldots,K)$ is the stress level of implementing ADT. Then, there will be K^L combined modes of stresses. Implementing ADT with all the K^L combinations is very expensive. Uniform orthogonal test theory can be used to choose the combined mode form the K^L combined modes scientifically and reasonably. According to this theory, test times is equal to the number of stress levels [11].

Take a test with two stresses and three levels for example, S_1 denotes stress 1, the stress levels are (S_{11}, S_{12}, S_{13}), S_2 denotes stress 2, the stress levels are (S_{21}, S_{22}, S_{23}), $S_{l0} < S_{l1} < S_{l2} < S_{l3} < S_{l\max}$ $(l = 1, 2)$. There will be 3^2 stress combinations. While according to the table of uniform orthogonal design $UL_3(3^2)$ (see Table 14.1). There are only three trials need to be implemented, and the stress

Table 14.1 $UL_3(3^2)$

Testing number	Stress factor A	Stress factor B
1	1	3
2	2	1
3	3	2

combinations are $\{(S_{11}, S_{23}), (S_{12}, S_{21}), (S_{13}, S_{22})\}$. Test profile is shown in Fig. 14.1.

It can ensure the uniform dispersion of the test factors and benefit the analysis after test using uniform orthogonal design. Assume that the total sample size is n, the total testing time is t, on the kth combination of stresses the proportion of sample size is p_k(then sample size $n_k = n \cdot p_k$) and the proportion of testing time is r_k (then testing time $t_k = t \cdot r_k$) ($k = 1,\ldots,K$). Let Δt denotes the interval of performance inspection, then the number of inspection on the kth combination of stresses $M_k = t_k/\Delta t$.

14.3 Accelerated Degradation Model

Model Assumptions

1. Degradation process is monotonic, i.e., the degradation damage can not reverse;
2. Degradation modes will not change with stress;
3. The remaining life of product depends only on the cumulative damage fraction it has happened and the current stress level;
4. Degradation process of the product follows the drift Brownian motion with drift $\mu > 0$ and dispersion $\sigma > 0$, which remains constant.

The degradation process $Y(t)$ can be described by the following drift Brownian motion model 0:

$$Y(t) = \sigma B(t) + d(s) \cdot t + y_0, \tag{14.1}$$

where, $B(t)$ is a standard Brownian motion on $[0, \infty)$, $B(t) \sim N(0, t)$; y_0 is the initial degradation level at time zero; σ is the dispersion and describes unit-to-unit variability and variability due to operating and environmental conditions, so it is assumed not to change with stress; $d(s)$ is the drift and commonly known as

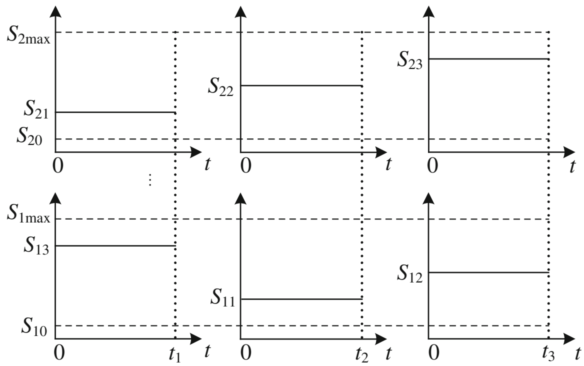


Fig. 14.1 Profile of CSADT with multiple stresses

degradation rate in industrial engineering and regarded as the function of stress conditions, i.e., drift parameter is an accelerated model.

For a given critical threshold C, the lifetime of the product follows the inverse Gaussian distribution with *p.d.f.* shown as follows,

$$f(t, C) = \frac{C}{\sigma\sqrt{2\pi t^3}} \exp\left\{-\frac{[C - d(s) \cdot t - y_0]^2}{2\sigma^2 t}\right\}, \tag{14.2}$$

Its reliability function is,

$$R(t) = \Phi\left[\frac{C - y_0 - d(s) \cdot t}{\sigma\sqrt{t}}\right] - \exp\left(\frac{2d(s) \cdot (C - y_0)}{\sigma^2}\right)\Phi\left[-\frac{C - y_0 + d(s) \cdot t}{\sigma\sqrt{t}}\right]. \tag{14.3}$$

For ADT with multiple stresses, accelerated model describing the relationship between multiple stresses and performance degradation of product should be used. In this paper we take two stresses for example, assume that there are some products mainly affected by temperature and voltage, so we suppose $d(s)$ follows the generalized Eyring model and there is no interaction between these stresses:

$$d(S) = \exp(A + B/T + C \cdot \ln V), \tag{14.4}$$

where A, B, C are constants; T is the absolute temperature with unit K; V is the voltage with unit v.

14.4 Optimization Model

14.4.1 Objective Function

We assume that there is no failure due to degradation. During the CSADT with L stresses, and K levels for each stress, there are K combined modes according to the above analysis, and the time of the jth inspection of the ith sample on the kth combination of stresses ($k = 1,\ldots,K$, $i = 1,\ldots,n$, $j = 1,\ldots M_k$) is t_{kij}, and the corresponding inspection result is y_{kij}. The Brownian motion is a Gaussian process, so the increment ΔD ($\Delta D = y_{kij} - y_{ki(j-1)}$) over Δt is independent and normally distributed with mean $d(s)\Delta t$ and variance σ^2 Δt [12]. The *p.d.f.* of independent increment is,

$$f(\Delta D) = \frac{1}{\sigma\sqrt{2\pi\Delta t}} \exp\left\{-\frac{[\Delta D - d(S) \cdot \Delta t]^2}{2\sigma^2 \cdot \Delta t}\right\}, \tag{14.5}$$

The log-likelihood function is,

$$\ln L \propto -\frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\sum_{j=1}^{M_k}\left\{\left[\ln(2\pi\Delta t)+\ln(\sigma^2)\right]+\frac{\left[\Delta D_{kij}-d_k(S)\cdot\Delta t\right]^2}{\sigma^2\Delta t}\right\}, \quad (14.6)$$

where the unknown parameters are $\theta = (A, B, C, \sigma)$.

Then the MLE of $\boldsymbol{\theta}$ can be obtained by numerical computation. Hence, the corresponding estimation of the pth quantile lifetime of products at use condition is $\hat{\xi}_p = \hat{R}^{-1}(p)$. When $n \to \infty$, $R(\xi_p)$ has an asymptotically normal distribution with mean $R\left(\hat{\xi}_p\right)$, and variance $\mathbf{h}^{\mathrm{T}}\mathbf{F}^{-1}(\boldsymbol{\theta})\mathbf{h}$. $\mathbf{F}(\boldsymbol{\theta})$ is the fisher information matrix for $\boldsymbol{\theta}$.

$$AsVar(R(\xi_p)) = \mathbf{h}^T\mathbf{F}^{-1}(\theta)\mathbf{h}. \quad (14.7)$$

$$\begin{aligned}\mathbf{F}(\theta) &= \frac{nt}{\sigma^2} \\ &\cdot\begin{bmatrix} \sum_{k=1}^{K} p_k r_k (d_k)^2 & \sum_{k=1}^{K} p_k r_k \frac{(d_k)^2}{T_k} & \sum_{k=1}^{K} p_k r_k \ln V_k\cdot(d_k)^2 & 0 \\ & \sum_{k=1}^{K} p_k r_k \frac{(d_k)^2}{T_k^2} & \sum_{k=1}^{K} p_k r_k \frac{\ln V\cdot(d_k)^2}{T_k} & 0 \\ & & \sum_{k=1}^{K} p_k r_k (\ln V_k)^2\cdot(d_k)^2 & 0 \\ \text{symmetrical} & & & \sum_{k=1}^{K} p_k r_k \Big/ 2\sigma^2\Delta t \end{bmatrix} \\ &= \frac{nt}{\sigma^2}\cdot\mathbf{I}(\theta).\end{aligned} \quad (14.8)$$

$$\begin{aligned}\mathbf{h}^T &= \left(\frac{\partial R(\xi_p)}{\partial A}, \frac{\partial R(\xi_p)}{\partial B}, \frac{\partial R(\xi_p)}{\partial C}, \frac{\partial R(\xi_p)}{\partial \sigma^2}\right) \\ \frac{\partial R(\xi_p)}{\partial \theta_i} &= \lim_{\Delta\theta_i\to 0}\frac{R(\xi_p,\theta_i+\Delta\theta_i)-R(\xi_p,\theta_i)}{\Delta\theta_i}\end{aligned}. \quad (14.9)$$

14.4.2 Constraint Conditions

Constraint conditions can be divided into two parts: experimental cost constraint and decision variables' actual ranges constraint.

The total experimental cost consists of two parts:

1. Cost of operation: It includes the operator's salary, power bills, depreciation of the testing equipment, and measure equipment, etc. It is related to the testing time and is $t{\cdot}C_o$, where C_o denotes the unit cost of operation;
2. Cost of testing devices: This is related to the number of testing devices and is $n{\cdot}C_d$;

Therefore, the experimental cost constraint is $t{\cdot}C_o + n{\cdot}C_d \le C_t$. C_t is the predetermined budget.

Actual ranges constraint of decision variables are as follows,

1. $S_{l0} < S_{l1} < S_{l2} < \ldots < S_{lK} \le S_{l\max}$, and $d_1(S) < d_2(S) < \ldots < d_K(S)$. For providing more accuracy of extrapolation for the degradation rate under normal condition, we should ensure that the degradation rate under the kth combination of stresses should be lower than under the $(k+1)$th combination and make a greater "span" between the degradation rates between two stress combinations.
2. $1 > r_1 \ge r_2 \ge \ldots \ge r_K > 0, \quad \sum_{k=1}^{K} r_k = 1, t_k = t \times r_k$. Assume that degradation modes will not change with stress, t_i under $d_i(S)$ should be set longer than t_j under $d_j(S)$, when $d_i(S) < d_j(S)$.
3. $p_k \ge 3/n, \sum_{k=1}^{K} p_k = 1, \quad (\mathrm{k} = 1{,}2{,}\ldots{,}\mathrm{K})$. Statistically n_k should not be less than 3.

Optimization problem can be obtained as follows. The decision variables are t, n, S_{lk}, p_k, and r_k.

$$\begin{array}{ll} \min & AsVar(R(\xi_p)) = \mathbf{h}^T\mathbf{F}^{-1}(\theta)\mathbf{h} = \frac{\sigma^2}{nt} \cdot \mathbf{h}^T\mathbf{I}^{-1}(\theta)\mathbf{h} \\ s.t. & t \cdot C_o + n \cdot C_d \le C_t \\ & S_{l0} < S_{l1} < S_{l2} < \cdots < S_{lK} \le S_{l\max}(l = 1, \cdots, L) \\ & d_1(S) < d_2(S) < \cdots < d_K(S) \\ & 1 > r_1 \ge r_2 \ge \ldots \ge r_K > 0, \sum_{k=1}^{K} r_k = 1 \\ & p_k \ge 3/n, \sum_{k=1}^{K} p_k = 1.(k = 1, \cdots, K) \end{array}. \tag{14.10}$$

We can see form the above equation that $AsVar(R(\xi_p))$ is a function of n, t, Δt, S_{lk}, pk, and r_k. For minimizing $AsVar(R(\xi_p))$, we should add sample size n and prolong testing time t.

14.4.3 Algorithm of the Optimization Problem

Figure 14.2 shows the algorithm flow of this optimization problem. Firstly, according to the objective function and experimental cost constraint we determine the solution spaces for n and t. Second, according to the actual situation the search step for each variable should be defined. For example, let the step of temperature T is 5 °C, the step of voltage V is 1v, the step of r_k is 0.1, and the step of p_k is $1/n$, then, according to these variables' actual ranges constraint, we find their solution spaces. Finally, we build a scheme set for the optimization problem including the solution spaces for these decision variables, and compute $AsVar(R(\xi_p))$ for each scheme in the set, then find the minimal value and the corresponding scheme is the optimal solution.

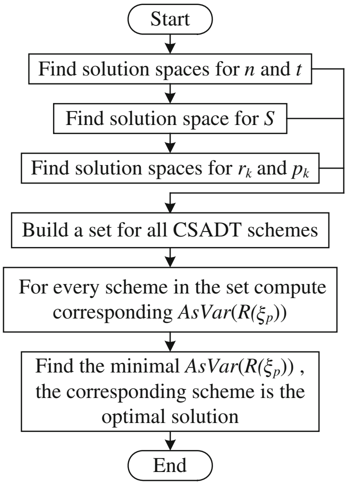


Fig. 14.2 Algorithm flow

14.5 Simulation Study

We set $(C_o, C_d, C_t) = (0.2, 2.5, 100) \times 103$ RMB. Form Eq. (14.10), we can obtain the optimal result: n = 20, and t = 250.

Based on the design information of products and limit analysis of stresses, the highest temperature and voltage are 110 °C and 15 v, and normal conditions are 40 °C and 5 v respectively. Model parameters $(A, B, C, \sigma) = (5, -7543, 0.4, 0.002)$. We set stress level number K is 3, 4, 5 respectively, the optimal results are shown in Table 14.2. The 3-D figure of accelerated model and the degradation rates of stress combinations in Table 14.2 are described in Fig. 14.3. We can see that the degradation rates satisfy the constraint condition.

Table 14.2 Optimal results for CSADT

K	$S_1,\ldots,S_K$	$p_1,\ldots,p_K$	$n_1,\ldots,n_K$	$r_1,\ldots,r_K$	$t_1,\ldots,t_K$	$AsVar(R(\xi_p))$ %	
3	T_k	75, 105, 110	0.15, 0.65, 0.2	3, 13, 4	0.4, 0.4, 0.2	100, 100, 50	0.034
	V_k	15, 10, 14					
4	T_k	80, 100, 105, 110	0.2, 0.15, 0.5, 0.15	3, 3, 10, 3	0.5, 0.2, 0.2, 0.1	125, 50, 50, 25	0.063
	V_k	13, 15, 10, 14					
5	T_k	80, 95, 100, 105, 110	0.15, 0.15, 0.15, 0.35, 0.2	3, 3, 3, 7, 4	0.2, 0.2, 0.2, 0.2, 0.2	50, 50, 50, 50, 50	0.081
	V_k	11, 15, 12, 10, 14					

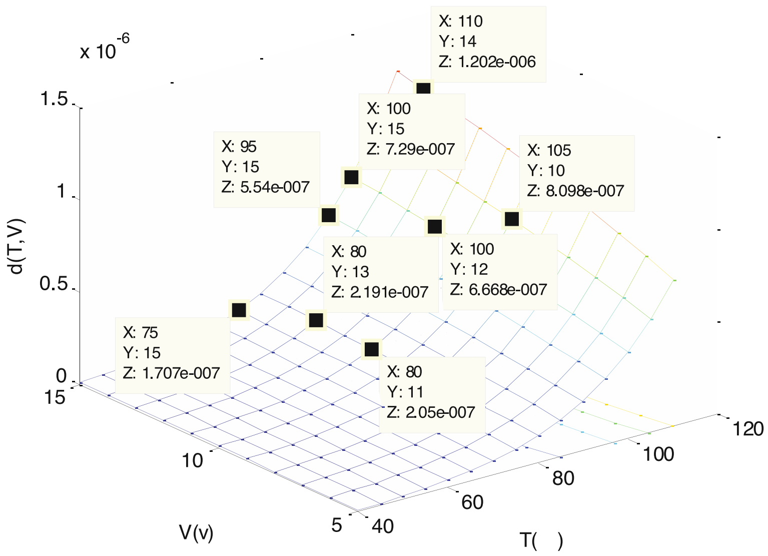


Fig. 14.3 3-D figure of accelerated model

We can see form Table 14.2 that,

1. $AsVar(R(\xi_p))$ is less than 0.1 %, so it can reduce the error of estimation efficiently using the scheme in the table.
2. The precision of estimation is the highest when $K = 3$. K does not affect the accuracy of estimation significantly. In actual, K should be set bigger than 3 for the precision and validity of extrapolation.
3. The search step of every decision variable affects the scheme set. The optimal result will be more accurate if the search step decrease, but the operation time of the algorithm will be longer. So the search step should be set reasonably according to the actual situation

14.6 Conclusions

In this paper, we derive the optimal test plans for CSADT with multiple stresses in which the degradation process can be appropriately described by drift Brownian motion. We derive the optimal test plans by using the criterion of minimizing the asymptotic variance of the estimation of the reliability of the pth quantile of product's lifetime under use condition, subject to the constraint that the total experimental cost does not exceed a predetermined budget. The optimal plan can give stress combination, stress levels, the sample size, and testing time under each stress combination.

From the simulation examples, we know that it can reduce the error of estimation efficiently using the optimal test plan and the number of stress level K should be set bigger than 3 for the precision and validity of extrapolation.

Uniform orthogonal test theory is introduced to determine the combined mode of diffident stresses in CSADT. It can reduce test times scientifically, and make the test factors have uniform dispersion.

References

1. Boulanger M, Escobar LA (1994) Experimental design for a class of accelerated degradation tests. Technometrics 36(3):260–272
2. Li Q (2002) Accelerated degradation test planning and optimization. PhD Thesis, The University of Arizona, Tucson, Arizona
3. Park SJ, Yum BJ, Balamurali S (2004) Optimal design of step-stress degradation tests in the case of destructive measurement. Qual Technol Quant Manage 1(1):105–124
4. Polavarapu I, Okogbaa G (2005) An interval estimate of mean-time-to-failure for a product with reciprocal weibull degradation failure rate. In: Proceedings annual reliability and maintainability symposium. IEEE, Alexandria, p 261–264
5. Yu HF (2006) Designing an accelerated degradation experiment with a reciprocal Weibull degradation rate. J Stat Plann Infer 136(1):282–297
6. Tseng ST, Balakrishnan N, Tsai CC (2009) Optimal step-stress accelerated degradation test plan for gamma degradation processes. Trans Reliab 58(4):611–618
7. Wang YS (2008) Research of simulation-based optimal design for accelerated testing. PhD Thesis, National University of Defense Technology, Changsha (in Chinese)
8. Li XY, Jiang TM (2009) Optimal design for step-stress accelerated degradation testing with competing failure modes. In: Proceedings annual reliability and maintainability symposium, p 64–68
9. Ge ZZ, Li XY, Jiang TM, Huang TT (2011) Optimal design for step-stress accelerated degradation testing based on D-optimality. In: Proceedings annual reliability and maintainability symposium, Orlando p 348–353
10. Pan ZQ, Zhou JL, Peng BH (2009) Design of accelerated degradation tests with several stresses based on Wiener process. Syst Eng Theory Pract 29(8):64–71 (in Chinese)
11. Kaitai F, Changxing M (2001) Orthogonal and uniform experimental design. Science Press, Beijing (in Chinese). http://www.math.hkbu.edu.hk/UniformDesign/
12. Chhikara RS, Folks JL (1989) The inverse Gaussian distribution. Marcel Dekker, New York

Chapter 15
Group Maintenance Scheduling: A Case Study for a Pipeline Network

F. Li, L. Ma, Y. Sun and J. Mathew

Abstract This paper presents a group maintenance scheduling case study for a water distribution network. This water pipeline network presents the challenge of maintaining aging pipelines with the associated increases in annual maintenance costs. The case study focuses on developing an effective pipeline replacement planning for the water utility. Replacement planning involves large capital commitment and can be difficult as it needs to balance various replacement needs under limited budgets. A Maintenance Grouping Optimization (MGO) model based on a modified genetic algorithm was utilized to develop an optimum group maintenance schedule over a 20 year cycle. An adjacent geographical distribution of pipelines was used as a grouping criterion to control the searching space of the MGO model through a Judgment Matrix. Based on the optimum group maintenance schedule, the total cost was effectively reduced compared with the schedules without grouping maintenance jobs. This optimum result can be used as a guidance to optimize the current maintenance plan for the water utility.

F. Li (✉) · L. Ma · Y. Sun · J. Mathew
CRC for Infrastructure and Engineering Asset Management, Faculty of Built Environment & Engineering, Queensland University of Technology, Brisbane 4001 QLD, Australia
e-mail: fengfeng.li@qut.edu.au

L. Ma
e-mail: l.ma@qut.edu.au

Y. Sun
e-mail: y3.sun@qut.edu.au

J. Mathew
e-mail: j.mathew@qut.edu.au

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_15,

15.1 Introduction

Previously, a Maintenance Grouping Model has been proposed by the authors [1] to consider group scheduling for pipeline maintenance. This paper presents a case study for the application of this model in the maintenance scheduling of a water distribution network in Australia. Some of the parts of the pipelines were aging which presented a challenge to the maintainers as the number of failures of these pipelines was rising over the years which cost millions of dollars to maintain annually. Pipelines can be maintained individually or in groups. This case study examines how these maintenance schedules can be grouped together to reduce total cost, and considers the trade off between maintenance cost and reducing the number of failures in the network using life cycle cost analysis.

The paper reports on the benefits of a group maintenance scheduling program for a water distribution network based on a Maintenance Grouping Optimization model, which makes for an optimum solution for grouping maintenance schedules at targeted planning horizons over the next 20 years.

The paper begins with a literature review of the core theories and existing research relevant to pipeline maintenance decision planning in Sect. 15.2, followed by an introduction of the new model for group maintenance scheduling. An overview of the water pipeline network is introduced in Sect. 15.3. In Sect. 15.4, a 20 year group maintenance schedule is presented for the water pipeline network utilizing the new model introduced in Sect. 15.2. The analysis of the results is presented in Sect. 15.4. A summary of the work performed to date and a proposal for future work is presented in Sect. 15.5.

15.2 Theory and Literature

Whether to replace or maintain a pipeline is an extremely significant and difficult decision for utility owners. A good maintenance decision is about maintaining the right pipe, at the right time, and using the right maintenance strategy and technology. Literature on maintenance planning for water distributed network offers interesting approaches.

Based on a statistical analysis of pipe lifetimes, a software KANEW [2] was developed, which deals with system wide renewal decisions for different water main categories. This water mains replacement scheduling model can schedule replacement in any time intervals, with the expenditure in each period constrained by available funds. Engelhardt [3] developed a method by allowing the replacements to be scheduled over a 20 year period, which was split into four five-year time periods with the expenditure in each period constrained by available funds considering various time-dependent parameters, such as increases in demand. Sægrov presented a decision support system CARE-W [4], which allowed selecting and scheduling rehabilitation jobs taking into account deterioration. This software provided a hydraulic modeling to access pipeline reliability of the network. Based on risk calculation, PARMS-PRIORITY [5] was developed that included pipeline failure prediction, cost assessment, data exploration, and scenario evaluation.

A non-homogeneous Poisson model I-WARP [6] was used to model the probability of breakage in individual water pipelines. It allowed for the consideration of three classes of covariates, pipe-dependent, time-dependent, and pipe and time-dependent. A renewal scheduling model [7] was developed for water main renewal planning, which took into account life cycle costs and associated savings due to reduced mobilization costs by setting a contiguity discount. However, it only considers two pre-determined situations, i.e., that two pipes share the same node and both are replaced in the same year.

Most of maintenance decision models in previous research only considered individual pipes. However, from the water utilities owner's point of view, group maintenance scheduling can certainly reduce the cost of replacement. In current practice, replacement planners usually have no choice but select two or three pipes manually with ambiguous criteria to group as one replacement job. This practice is not a viable solution for pipe grouping and is not cost effective. To address this issue, a novel concept of group maintenance scheduling was introduced by Li [1]. The proposed Maintenance Grouping Optimization (*MGO*) model was shown to assist distributed pipeline assets owners and operators to make group scheduling decisions for replacement of water pipelines. They created a Judgment Matrix considering three grouping criteria (adjacent geographically distribution, identical replacement machinery, and similar service interruption areas). This Judgment Matrix corresponded to various combinations of pipe replacement schedules. A modified cost model was developed to calculate the total cost of each pipeline at each year. A modified Genetic Algorithm model was proposed to deal with the scalability of the vast combinatorial solution space for finding the optimum group maintenance scheduling option, with the objective of minimizing total cost.

This case study started from three cost models and adopted the Non-homogeneous Poisson based model to predict the probability of breakages and the per-analysis (calculating the total cost) to filter the pipes that will not be replaced during the 20 year planning horizon. The proposed group maintenance scheduling model was subsequently utilized with one of the three grouping criteria (adjacent geographically distributed) to find the optimum maintenance schedule. The influences of customer interruption of the group maintenance schedule were not included in this work.

15.3 Overview of the Pipeline Network

15.3.1 Overview of the Water Distributed Network

This water distributed network services a community in Australia, and comprises 66,405 pipes (total length of 3,640 km, at an average length of pipe of 54.79 m),

Table 15.1 Overview of the water distributed network

Item	Description
Pipe's diameter	20, 32, 40, 45, 50, 63, 75, 80, 90, 100, 110, 150, 200, 220, 250, 300, 375, 411, 450, 500, 510, 525, 565, 590, 600, 660, 700, 750, 800, 850, 900, 915, 960, 965, 1,050, 1,125, 1,290, 1,440 mm
Pipe's material (11 types)	Asbestos Cement (AC), Cast Iron Cement Lined (CICL), Concrete (CONC), Copper (CU), Fiber Reinforced Pipe (FRR), Galvanized Steel (GAL), Glass Reinforced Plastic (GRP), Glass Reinforced Plastic (HOBAS), Ductile Iron Cement Lined (DICL), Mild Steel Concrete Lined (MSCL), Unplasticized Poly Vinyl Chloride (UPVC)
Pipe's length	From 0.1 to 1363.9 m
Zone area types	RURAL (RUR), URBAN (URB), HIGH DENSITY URBAN (HDU), CBD
Population	More than 500,000 People

with diameters from 20 to 1,440 mm, in 11 different materials, installed between 1937 and 2010. It services a population of more than 500,000 inhabitants through 9 different workstations. An overview of the network is presented in Table 15.1.

15.3.2 Age Profile of the Water Pipeline Network

The oldest water pipes in the network dates back to 1937. Around 102 km of the total length of pipes now in operation were installed before 1960. Almost all pipes were made from concrete or cement before 1960. The years 1961 to 1990, saw a significant increase in installations, and nearly half of the total number of pipes (3/5 of total length) were constructed in this period. The construction history is shown in Fig. 15.1. The most commonly used materials were ductile iron, gray

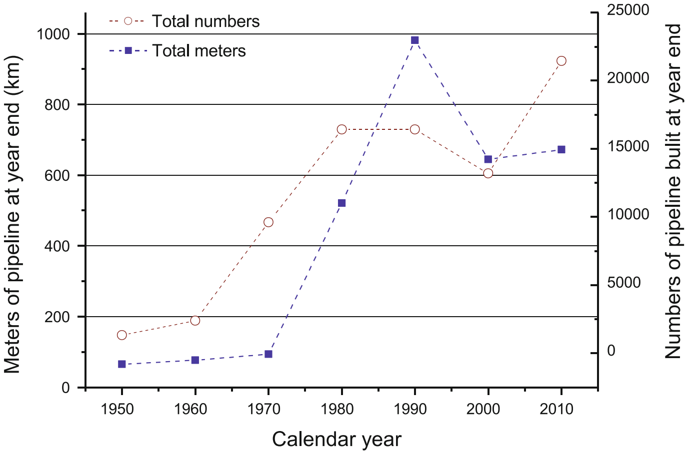


Fig. 15.1 Construction history

cast iron, and mild steel. In 1960, the first PVC pipes were constructed and have become the preferred material for replacements and expansions after 1970s.

The average age of the whole network is 21 years, but the age of 1,356 pipes (2 % of total network) exceed 60 years, i.e., 70 km in total.

15.3.3 Pipeline Repair History

The water company recorded 3,128 pipe repairs during 2000–2010, with the structure of the work order as follows: pipe ID, depth (>1.5 and <1.5 m), repair start and finish date and time, and repair cost. Naturally some of these data were found to be missing for various reasons while 2,526 sets of valid data were filtered for analysis. Over 10 years, the water company conducted 3,823 repair jobs for unexpected breaks, which means several pipes broke twice or more, and cost AUD$4 million. Figure 15.2 shows that the repair cost correlated with the number of breaks which rose from 2000 to 2010, and decreased slightly in 2005, and then peaked at AU$0.86 million in 2010. The drop in the number of breaks in 2005 remains unexplained and is being analyzed further at this time.

The 2,526 records of pipeline repair corresponded to five different pipe materials, which were Asbestos Cement (AC), Cast Iron Cement Lined (CICL), Ductile Iron Cement Lined (DICL), Mild Steel Concrete Lined (MSCL), and Unplasticized Poly Vinyl Chloride (UPVC). These pipes were installed from 1937 to 1990, and were generally near their end of life (60–80 years). A statistical analysis was conducted to analyze the relationships between the repair cost and materials as well as repair cost and diameter. Two box and whisker plots are illustrated in Fig. 15.4 to show different materials and diameters of repair cost data through the

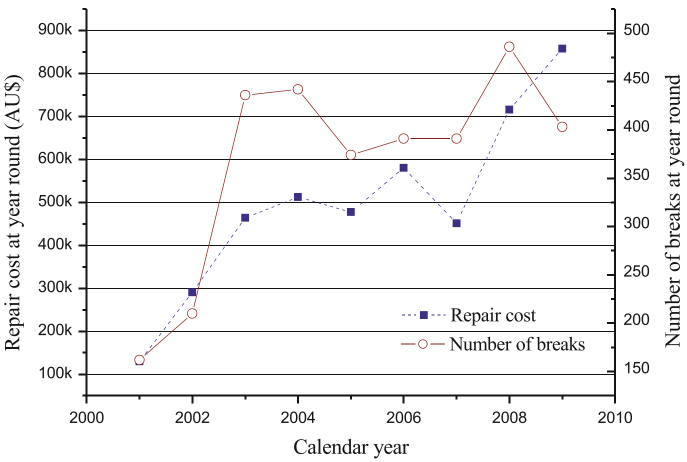


Fig. 15.2 Repair history from 2000 to 2010

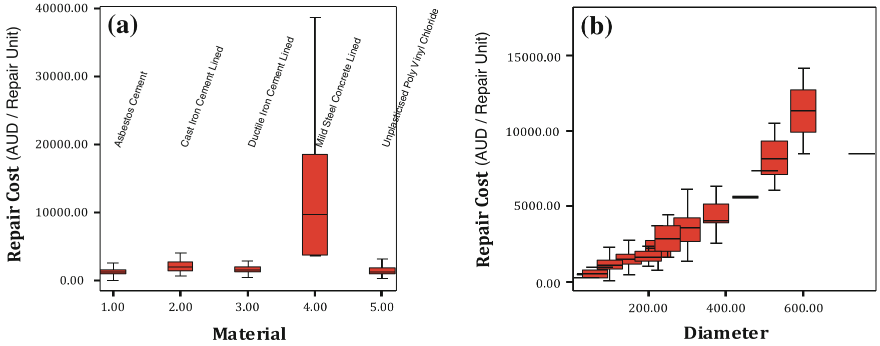


Fig. 15.3 Repair cost with different materials and diameters. **a** Repair cost w/different materials. **b** Repair cost w/different diameters

smallest cost, lower quartile, mean value, upper quartile, and the largest cost observation. Figure 15.3a illustrates that the repair cost shows dramatic differences between MSCL and the other materials. Three reasons caused the high price of MSCL pipes' repair cost: (1) the price of this material on its own was extremely higher than the other materials; (2) the repair method and procedure utilized in MSCL pipes were more complicated that the other pipes; (3) larger diameters were used in the MSCL pipes (>300 mm). Moreover, the other materials showed a similar repair cost in this case, so that one can treat these four materials as a group, when the impact of material is taken into account.

Figure 15.3b shows that the repair cost increased with the increase in diameter as shown in [8, 9]. Based on the repair data in this case, the relationship between the increase in the repair cost and pipe diameter was found to be nonlinear.

15.4 Group Maintenance Scheduling

15.4.1 Repair Cost

Based on the 2,526 records of pipeline repair data, a regression model [10] was used and trained to calculate repair cost of each pipe i, based on the Eq. (15.1):

$$C_i^{rep} = a + b \cdot D_i^c + d \cdot u_i^e + f \cdot D_i \cdot u_i. \tag{15.1}$$

D_i indicates the diameter of each pipe i, and the decision variable u_i is the depth of the pipe buried, where "1" indicates the depth <= 1.5 m, and "2" indicates the depth >1.5 m. *a, b, c, d, e,* and f are coefficients, which are estimated by using the regression model. In this case, u_i is independent of D_i, therefore, parameter f is equal to 0. The coefficients of the repair cost equation with five different materials of pipes are given in Table 15.2.

Table 15.2 Parameters for repair cost equation

Material of pipe	a	b	c	d	e	R^2
AC, CICL, DICL, UPVC	877.201	0.066	1.849	0.233	−0.146	0.99
MSCL	659.143	0.023	2.063	−0.002	19.399	0.89

15.4.2 Replacement Cost

Replacement costs are affected by the length, diameter, replacement technology, and location of each pipe. The replacement cost function of each pipe i, C_i^{repl}, [1], is shown in Eq. (15.2), which contains three components, the cost related to length Cr, the machinery cost M_i, and the travel cost Cv_i:

$$C_i^{repl} = \frac{M_i}{\sum\limits_{\forall i \in U} l_i} \cdot l_i + \frac{l_i \cdot \sum\limits_{\forall i \in U} (Cr_i \cdot l_i)}{\sum\limits_{\forall i \in U} l_i} + Cv_i \frac{d_i^2}{\sum\limits_{\forall i \in U} d_i}, \tag{15.2}$$

where pipe i belongs to group U, d_i is the travel distance from work station to pipe i. The length related cost Cr_i is correlated to various diameters and materials, which shows in Table 15.3. The information in Table 15.3 is based on the historical contract payments for the last year for water main replacements.

The factors affecting machinery cost remain unexplained in the current research. Therefore, it was assumed that the machinery cost is generally affected by the materials. Considering the analysis of repair cost, the results showed dramatic differences between MSCL and other materials. According to the case study made by Li, and considering the dramatic differences of the repair cost between the results of the MSCL and the other materials, the machinery cost was taken as

Table 15.3 Water pipes replacement cost (Cr_i)

Diameter (mm)	Substituted material	Cr_i (AU$/m)	Diameter (mm)	Substituted material	Cr_i (AU$/m)
90	PVC	$98	900	DICL	$2,575
100	PVC	$104	960	MSCL	$2,901
150	PVC	$168	1,000	MSCL	$3,046
200	PVC	$229	1,050	MSCL	$3,152
225	PVC	$254	1,085	MSCL	$3,326
250	PVC	$273	1,200	MSCL	$3,748
300	DICL	$461	1,290	MSCL	$4,007
375	DICL	$654	1,350	MSCL	$4,350
450	DICL	$777	1,500	MSCL	$4,769
500	DICL	$946	1,650	MSCL	$5,316
525	DICL	$1,020	1,800	MSCL	$5,765
600	DICL	$1,240	1,950	MSCL	$6,324
750	DICL	$1,759	2,159	MSCL	$6,792

$M = AUD\$4{,}000/unit$ for MSCL and $M = AUD\$2{,}000/unit$ for other materials. The CV is distance-unit cost, which is a constant value, and is set at AUD$100/km in this case.

15.4.3 Total Cost

The total cost C^{tot} associated with pipe replacement is affected by the replacement cost C_i^{repl} and pipeline failures [7]. The cost of pipeline failure includes two parts, (1) direct cost, which contains repair cost C^{rep}, direct damage cost C^{dir} and water loss cost C^{wat}, and (2) indirect cost, which comprises indirect damage cost C^{indir}, customer service interruption cost C^{inp}, and social cost C^{soc}. In the current case, the authors considered the repair cost as an only component of failure cost, for it was impossible to calculate damage cost, water loss cost, customer service interruption cost, and social cost. Therefore, the present value of the total cost associated with pipe i replaced at year t is given by:

$$C_{i,t}^{tot} = C_i^{repl} \Big/ (1+r)^t + \sum_{p=1}^{t} k_{i,p} [C_i^{rep} \cdot \frac{1}{(1+r)^p}], \tag{15.3}$$

where r is a discount rate which is equal to “0.02”, and $k_{i,p}$ is the probability of breakage in pth year.

15.4.4 The Objective Function

The objective, in this case, is to minimize the present value of total cost of water distribution network. This took the group maintenance scheduling of candidate pipes into account during the T (20 year) planning horizon. The minimum total cost of the network during horizon T C_{sys}^{tot}, is equal to the minimum cost of sum of the total cost of each pipe i in the selected tth year, $t \in (1, 2, \ldots T)$:

$$Min \quad C_{sys}^{tot} \quad = Min \quad \sum_{\forall t \in T} \sum_{i}^{N} C_{i,t}^{tot}\,. \tag{15.4}$$

Constraints

1. The total cost of the pipes replacement in the planning horizon T years must be less than the total budget B_T

$$\sum_{\forall t \in T} \sum_{i}^{N} C_{i,t}^{tot} \le B_T, \tag{15.5}$$

where the annual budget of the whole water distribution network for replacement was AUD$2 million. Therefore, the total budget of the whole water distributed network for replacement B_T is equal to $B_t \times T$, which was AUD$40 million for 20 years.

2. For all grouping options, the distance from pipe i to pipe j must be equal or smaller than the maximum distance, which can be set by users.

15.4.5 NHPP Analysis

An analysis of forecast failure probability was conducted to screen the target pipes from all pipes (66,405 pipes). A well proved non homogeneous Poisson-based model [6], which is capable of considering time-dependent covariates, was used to forecast the probability of breakages. It was assumed that the breaks at year t for an individual pipe i satisfied Poisson process with mean intensity $\lambda_{i,t}$. The probability of breakages $k_{i,t}$ is given by:

$$P(k_{i,t}) = \frac{\lambda_{i,t}^{k_{i,t}} \cdot e^{-\lambda_{i,t}}}{k_{i,t}!}, \tag{15.6}$$

$$\lambda_{i,t} = \exp\left[\alpha_0 + \theta \log\left(g_{i,t}\right) + \alpha \log\left(z^i\right) + \gamma q^{i,t}\right], \tag{15.7}$$

where $g_{i,t}$ is the age of pipe i at year t, z^i is the length of pipe i, and $q^{i,t}$ is the number of known previous failure breaks. α_0, θ, α, and γ are coefficients. The 2,526 records of pipe repairs were used for training the model to find the four coefficients, which are shown in Table 15.4.

The forecasted failure probabilities of each pipe i at each year t were calculated by using the prediction model. The forecasted failure probabilities of selected seven pipes over 20 year planning are shown in Fig. 15.4. Figure 15.5 illustrates the relationship between the failure probabilities and the number of pipes. Almost 80 % of total pipes' failure probabilities were lower than 0.1, and only a few pipes' failure probabilities were higher than 0.1.

To increase the computing efficiency, only the pipes whose forecasted failure probabilities were higher than 0.1 were considered for further analysis. If the failure probability of one pipe is lower than 0.1*, which* indicates that this pipe was unlikely to fail in the future years, then this pipe do not need to replace. According to this criterion, 7,191 (23 %) pipes were selected for the further analysis.

Table 15.4 Coefficients of NHPP model

Covariate	α_0	θ	α	γ
Coefficient	−3.846	0.54	0.75	0.32

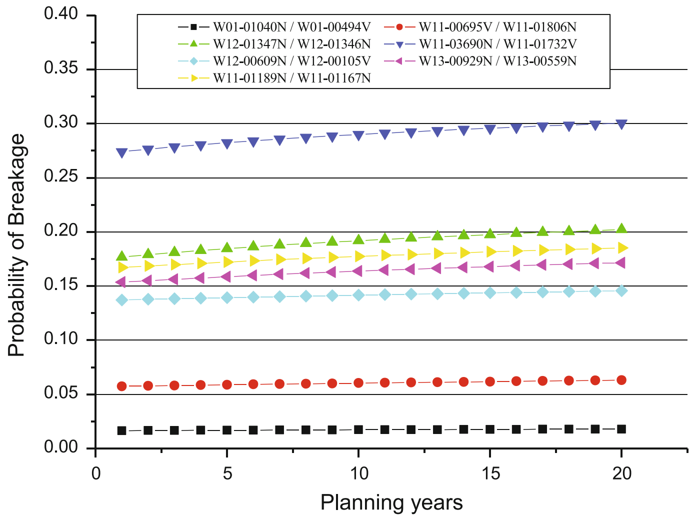


Fig. 15.4 Sample of forecasted failure probability

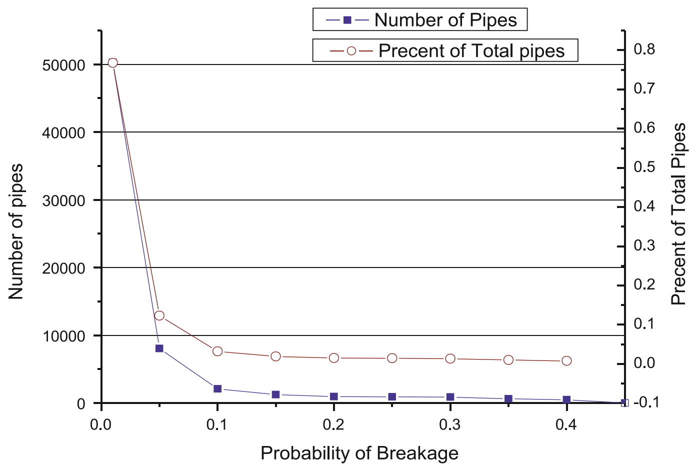


Fig. 15.5 Distribution of the failure probability with pipe numbers

15.4.6 Pre-analysis

Based on Eqs. (15.1–15.5), the total cost of each pipe i at each selected year t was calculated, without considering group maintenance schedules. 1,092 pipes were selected from the 7,191 pipes with high failure probabilities, whose minimum cost happened within the decision horizon $T = 20$. The details of selected pipes are listed in Table 15.5.

Table 15.5 Summary of the selected pipes

Item	Description
Diameter	40, 50, 63, 100, 150, 200, 220, 300, 375, 450, 500, 525, 600, 660, 750, 850, 960, 965 mm
Material (5 types)	Asbestos Cement (AC), Cast Iron Cement Lined (CICL), Concrete (CONC), Ductile Iron Cement Lined (DICL), Mild Steel Concrete Lined (MSCL)
Length	From 2.1 to 1,843 m

15.4.7 Judgment Matrix

Judgment Matrix was calculated using the following equations [1],

$$\Lambda = \begin{bmatrix} \varepsilon_{11} & \cdots & \varepsilon_{1j} & \cdots & \varepsilon_{1n} \\ \vdots & \ddots & \vdots & & \vdots \\ \varepsilon_{i1} & \cdots & \varepsilon_{ij} & \cdots & \varepsilon_{in} \\ \vdots & & \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \cdots & \varepsilon_{nj} & \cdots & \varepsilon_{nn} \end{bmatrix}, \tag{15.8}$$

where $\varepsilon_{ij} = \begin{cases} \gamma_{ij}, & \gamma_{ij} \leq \gamma^* \\ 0, & \text{otherwise} \end{cases}$, γ_{ij} = Geographic distance from pipe i to pipe j (km), and γ^*= Maximum geographic distance (km), $i, j, n \in N$, i, j, n are the indexes of pipes, and N is the total number of pipes in the network.

The maximum geographic distance γ^* is an input value, which can be determined by users. In this case, the maximum geographic distance γ^* was determined through calculating the distance between the nine workstations. The minimum distance among the nine workstations was equal to 4.3 km. It was assumed that one replacement job can only be done by only one work team from only one workstation. Therefore, the maximum geographic distance γ^* was equal to the half of the minimum distance among the nine workstations, which means γ^* was equal to 2.15 km.

The judgment matrix in this case was a "1092 by 1092" matrix with the rational numbers from "0" to "2.15 km". Figure 15.6 shows the judgment matrix, the *black* color indicates that the distance was beyond the "2.15 km"*, which* means that there was no grouping pipes with corresponding pipes, the colors from "white" to "gray" means the distance from "0" to "2.15 km".

15.4.8 Group Maintenance Scheduling

For group maintenance scheduling, the Grouping model presented in [1] was utilized to analyze the different grouping options. The Grouping model was based on a modified Genetic Algorithm (GA) model, and the parameters of that GA model was listed in Table 15.6.

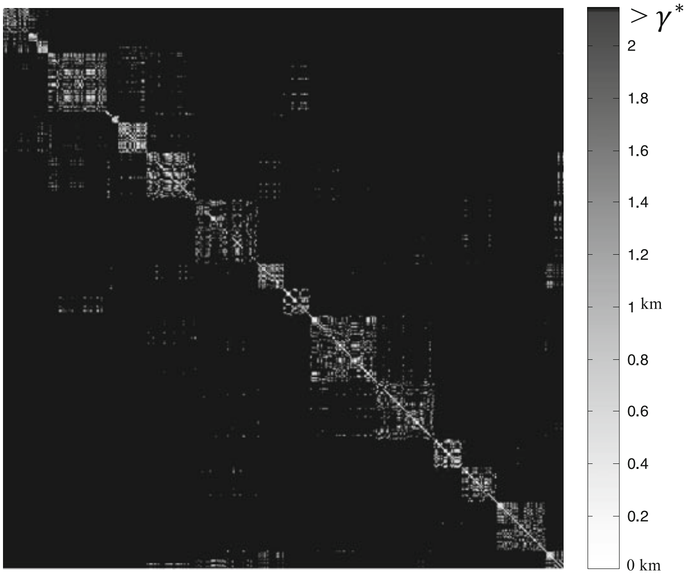


Fig. 15.6 Judgment matrix

Table 15.6 Parameters of the modified GA model

Number of individuals	300	Generation gap	0.9
Maximum generations	500	Maximum number of pipes grouped	5

The working time of one replacement job depends on many factors, i.e., pipe's length, location, diameter, material, and working efficiency of work team. Practically, one replacement job at least takes 4 h, therefore, we assumed that one group of jobs is not longer than 3 working days, which means that the maximum number of pipes in one group does not exceed five. According to the structure of the modified GA model, the genes number was decided by the maximum number of pipes in one group. Therefore, in this case study, the representation involved 6 'genes' for each pipe of the network, which shows in Fig. 15.7. Therefore, one chromosome had 6,552 (1,092 $\times$ 6) genes.

15.4.9 Results and Discussion

Using the MGO model, an optimum group maintenance schedule was planned for the planning horizon.

An example of first year planning is listed in Table 15.7. In the first year, 19 pipes will be replaced in 9 grouped replacement jobs. The total pipe lengths in

Pipe *ID*						…	...Pipe *ID*…						…	Pipe *ID*					
T	*Pipes Grouped*					…	*T*	*Pipes Grouped*					…	*T*	*Pipes Grouped*				
t	*ID*	*ID*	*ID*	*ID*	*ID*	…	*t*	*ID*	*ID*	*ID*	*ID*	*ID*	…	*t*	*ID*	*ID*	*ID*	*ID*	*ID*

Fig. 15.7 Chromosome representation

Table 15.7 Example of the first year planning Table 15.7

Job	Pipeline	Total length (m)
1	W01-00265 N/W01-01704 N	1,317
	W01-01414 N/W01-00154 V	
2	W01-00820 V/W01-01058 V	357
3	W01-00272 V/W01-00274 V	1,677
	W02-00195 N/W02-00174 V	
	W04-00592 N/W04-00697 V	
4	W04-00302 N/W04-00551 V	1,606
	W04-00588 N/W04-01795 N	
	W10-01237 N/W10-00434 V	
	W10-01502 N/W10-00888 N	
5	W10-00465 N/W10-00649 V	792
	W13-00607 V/W13-00608 V	
6	W07-00652 V/W07-00648 V	1,815
	W10-00467 V/W10-02169 N	
	W10-01847 N/W10-00446 V	
7	W10-01448 N/W05-00074 V	810
	W12-00098 V/W12-00094 V	
8	W15-00113 V/W15-00110 V	291
9	W12-01377 N/W12-01382 N	369

each job will be smaller than 2 km, which was assumed to be the maximum replacement length of one job.

Figure 15.8a shows the number of pipes planned to replace, the number of replacement jobs, and the total length of pipes planned to replace for the next 20 years. From the values in each year, it is noticed that there are significant differences among total number and total length of pipes replacement for each planning year, which breaks the convention that the total numbers and lengths of replacement pipes should slightly increase annually. Actually, the planning replacement jobs fluctuated during the planning horizon, and there was no clear regularity showed in the optimum replacement planning.

In Fig. 15.8b, the situation of replacement planning with considering group scheduling or not is compared. The annual expenses for both grouped and non-grouped scheduling are changed grammatically each year. The annual expenses of five different years for grouped scheduling is a little higher than the annual budget (AUD$2 million). On the contrary, the annual expenses of eight different years for non-grouped scheduling are higher than annual budget, especially in the first two

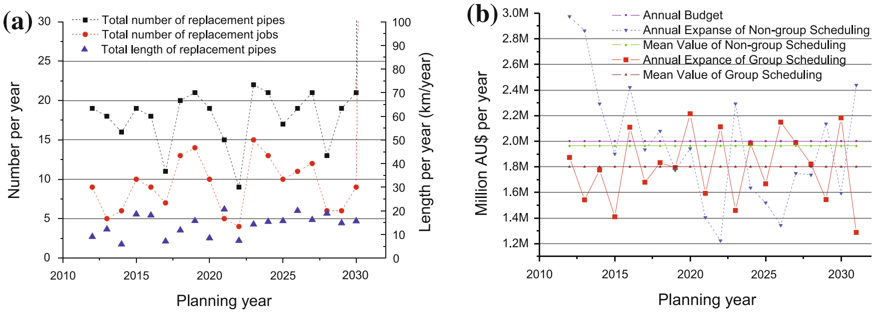


Fig. 15.8 Statistical information of the group scheduling. **a** Number, length and budget. **b** Comparison of annual expanse

years. The mean value for the total expenses of 20 years for the non-grouped scheduling is AUD$1.96 million each year, which is just below the budget. The mean value for the total expenses for the group scheduling is AUD$1.8 million each year, which has AUD$0.2 million and AUD$0.14 million cost saving compared with the annual budget and the non-grouped scheduling respectively. The total cost for the 20 years non-grouped scheduling is around AUD$39 million. On the contrary, the total cost for the 20 years group scheduling is around AUD$36 million, which is AU$4 million (10 %) and AU$3 million (7.5 %) cost reduction for the 20 planning years.

15.5 Conclusions

Optimum maintenance schedules are desirable for owners and operators of water pipeline networks to balance maintenance cost and the degradation of the network. To reduce total cost, maintenance schedules can be conducted in groups of pipes rather than individually. This paper validates that the Maintenance Grouping Optimization model based on a modified genetic algorithm is a good optimization model for techniques for managing pipeline maintenance, through developing an optimum group maintenance schedule for a water pipeline network over a 20 year cycle. The adjacent geographical distribution of pipelines can be used as a grouping criterion to control the searching space of the MGO model through a Judgment Matrix. Based on the optimum group maintenance schedule, the total cost can effectively come down compared with the schedules without grouping maintenance jobs. This model is a good technique for optimizing the current maintenance planning for the water utility.

Potential MGO model applied for other types of distributed assets including electricity distribution networks, railway networks, and road networks will be tested in the future research.

Acknowledgments We wish to thank Mr. Rex Mcbride and Mr. Bjorn Bluhe for providing useful comments and access to the data used in this research. This research was conducted within the CRC for Infrastructure and Engineering Asset Management, established and supported under the Australian Government's Cooperative Research Centres Program.

References

1. Li F et al. (2011) A grouping model for distributed pipeline assets maintenance decision. Proceedings of 2011 international conference on quality, reliability, risk, maintenance, and safety engineering. Xian, pp 627–632
2. Deb AK et al. (1998) Quantifying future rehabilitation and replacement needs of water mains. AWWA Research Foundation. p 156
3. Engelhardt MO (1998) Development of a strategy for the optimum replacement of water mains. Ph.D. Thesis. University of Adelaide: Department of Civil and Environmental Engineering, Adelaide. p 514
4. Saegrov S (2005) CARE-W: computer aided rehabilitation for water networks. IWA Publishing, London
5. Moglia M, Burn S, Meddings S (2006) Decision support system for water pipeline renewal prioritisation. ITcon Spec Issue Decis Support Syst Infrastruct Manage 11:237–256
6. Kleiner Y, Rajani BB (2010) I-WARP: individual water main renewal planner. Drinking Water Eng 3:71–77
7. Kleiner Y, Nafi A, Rajani B (2009) Planning renewal of water mains while considering deterioration, economies of scale and adjacent infrastructure. Water Sci Technol: Water Supply 10(6):89–906
8. Werey C, Llerena P, Nafi A (2008) Water pipe renewal using a multiobjective optimization approach. Can J Civ Eng 35(1):87–94
9. Christodoulou S, Charalambous C, Adamou A (2008) Rehabilitation and maintenance of water distribution network assets. Water Sci Technol: Water Supply 8(2):231–237
10. Robert MC et al (2002) Cost models for water supply distribution systems. J Water Resour Plan Manage 128(5):312–321

Chapter 16
Rail Wagon Bearings Health Management Based on Imperfect Acoustic Information

A. Ghasemi and M. R. Hodkiewicz

Abstract In this paper we develop and discuss a prognostics model to estimate the Mean Residual Life of Rail Wagon Bearings within certain confidence intervals. The prognostics model is constructed using a Proportional Hazards Model approach informed by imperfect data from a bearing acoustic monitoring system and related failure database. We have been able to predict failure within a defined maintenance planning window from the receipt of the latest acoustic condition monitoring information. We use the model to decide whether to replace a bearing or leave it until collection of the next condition monitoring indicators. The model is tested on a limited number of cases and demonstrates good predictive capability. Opportunities to improve the performance of the model are identified.

16.1 Introduction

This case study project develops a prognostics algorithm to determine Mean Residual Life (MRL) of Rail Wagon Bearings for use in maintenance planning in a mining organization. The main objectives are to (1) identify if an appropriate model can be built given available data, and (2) assess confidence in the model.

A. Ghasemi
Industrial Engineering Department, Dalhousie University,
5269 Morris Street, Halifax, NS B3J 2X4, Canada
e-mail: alireza.ghasemi@dal.ca

M. R. Hodkiewicz (✉)
Mechanical Engineering, University of Western Australia, M050,
Crawley, WA 6009, Australia
e-mail: melinda.hodkiewicz@uwa.edu.au

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_16,

This article also briefly describes the steps involved in developing the prognostics model, the technical and organizational assistance and barriers, data requirements, and factors that impact on confidence in the results.

The rail industry uses a variety of monitoring tools to assess the condition of the Rail wagon bearings. These include, but are not limited to, thermal (Hot Bearing Detection and Hot Wheel Detection), acoustic (Rail Bearing Acoustic Monitor), and weight-based techniques (Wheel Impact Detector). These provide condition indicators for diagnosis. Diagnostic interpretation and synthesis of the different indicators is predominantly intuitive and success is correlated with the experience of personnel familiar with the equipment. One of the main challenges is the number of bearings monitored. In this case study there are tens of thousands of Rail car wheel bearing assemblies in service.

Condition monitoring is the act of observing and collecting information concerning the degradation condition of equipment to schedule maintenance actions prior to failure. Data concerning one or more indicators of degradation are collected periodically to establish a diagnosis of the equipment's condition and a prognosis for future performance. A combination of human expertise, signal processing, and modeling is used through this process.

To apply any diagnosis and/or prognosis method to a real world problem, the models' parameters are usually estimated from the historical data. In this project a Proportional Hazards model (PH-Model) with imperfect information is used to determine a fundamental prognostic tool, the hazard function. According to the PH-model, the hazard function $h(t, Z(t))$ is represented by $h(t, Z(t)) = h_0(t)\psi(Z(t))$, where $Z(t)$ is the equipment's degradation state at time t. The most common baseline hazard function is the Weibull distribution function which represents the aging process, and the most used function for equipment degradation state is exponential in the form $\psi(Z(t)) = \exp(\gamma Z(t))$, where γ is the indicator's coefficient.

Realistically, information may contain noise due to errors of measurement, of interpretation, or due to the limited accuracy of the measurement's instruments, and it may not reveal the exact degradation state of the equipment. The information is, however, stochastically correlated with the degradation state. This category of condition monitoring is referred to as 'Indirect Monitoring' (Wang and Christer [1]), 'Partial Observation' (Makis and Jiang [2]), or Imperfect Observation by some others including Ghasemi et al. [3].

The ultimate output of this work is the Mean Residual Life (MRL) of the bearings after each pass-by with in a confidence interval that can be used to plan replacement of the bearing in a condition-based maintenance system. A wide range of parameter estimation methods for CBM models, incorporating the historical data, exist in the literature. In this work we will use the method introduced by Ghasemi et al. [4].

16.2 Approach

16.2.1 Proportional Hazards Models

The prognostics model is based on the PH-model proposed by Cox [5] while assumes that the condition monitoring is imperfect, i.e., at each observation moment, an indicator value, θ, of the underlying degradation state is available. Observations are collected at a constant (or a near constant) interval Δ. In this study, $Z(t)$ represents the degradation state of the equipment at time t, which will be used as the diagnostic covariate in the PH-model, and θ is a value from the set of all the possible indicator values $\Theta = \{1, \ldots, M\}$. This means that the set of indicator values is discretized into a finite set of M possible values. The equipment's condition is described as follows.

- The equipment has a finite, known number of degradation states, N. $J = \{1, \ldots, N\}$ is the set of all possible degradation states.
- The degradation state transition is modeled by a HMM. The transition matrix is $P = [p_{ij}]$, where $p_{ij} = \Pr(Z(t+\Delta)|Z(t) = i, T > t + \Delta)$ is the probability of going from state i to state j, $i,j \in J$ during one observation interval, knowing that the equipment does not fail before the end of the interval.
- The value of the indicator is stochastically related to the equipment's state through the observation probability matrix, $Q = [q_{j\theta}]$, $j \in J$, $\theta \in \Theta$. $q_{j\theta} = \Pr(\theta|Z(t) = j)$ is the probability of getting the indicator value θ, while the equipment is in degradation state j.
- The indicator is collected periodically at fixed or near fixed intervals Δ.
- Failure is not a degradation state. It is a non-working or dangerous condition of the equipment that can happen at any time, while the system is in any degradation state.

Ghasemi et al. [3] introduced a new state space, and transition rule for this problem. The transition rule includes all the observations from the last renewal moment, and provides a methodology to deal with unobservable degradation states by calculating the conditional probability of being in each degradation state. The conditional probability distribution of the equipment's degradation state at period k, is π^k defined as

$$\pi^k = \left\{\pi_i^k;\ 0 \le \pi_i^k \le 1 \text{ for } i = 1, \ldots, N, \sum_{i=1}^{N} \pi_i^k = 1\right\}; k = 0, 1, 2, \ldots, \quad (16.1)$$

where $\pi_i^0 = \begin{cases} 1 & i = 1 \\ 0 & \text{o.w.} \end{cases}$, meaning that the equipment is in its best possible state at period zero. After obtaining an indicator value θ via an inspection at an observation moment, the prior conditional probability π^k is updated to π^{k+1}. By using Bayes' formula, and knowing that the indicator θ has occurred at the $k+1$-st observation moment, $\pi_j^{k+1}(\theta)$ is determined as:

$$\pi_j^{k+1}(\theta) = \sum_{i=1}^{N} \pi_i^k p_{ij} q_{j\theta} \Bigg/ \sum_{i=1}^{N} \sum_{l=1}^{N} \pi_i^k p_{il} q_{l\theta}, \quad j = 1, \ldots, N. \tag{16.2}$$

16.2.2 Parameter Estimation

Let T be the lifetime of the equipment, which is an i.i.d., non-negative continuous random variable. $\theta(s) = \{\theta^1, \theta^2, \ldots, \theta^k\},\ s \le T; k = 1, 2, \ldots,\ k\Delta \le s$ is the history of the indicator values up to time s. Ghasemi et al. [6] maps the history of the indicator values up to time s into the state conditional probability distribution history $\pi(s) = \{\pi^1, \pi^2, \ldots, \pi^k\},\ s \le T,\ k = 1, 2, \ldots,\ k\Delta \le s,$ where the elements of $\pi(s)$ are calculated by (16.1), and (16.2) at corresponding observation moments, when a new indicator value is available.

By assuming that the equipment's unobservable degradation state transition follows a stationary Markov Process, the transition probability at time $t = k\Delta$, from state i to state j, knowing that the equipment has survived at least until the next observation moment is: $p_{ij}(k) = p_{ij} = \Pr(Z_{k+1} = j | Z_k = i, T > (k+1)\,\Delta),\ k = 1, 2, 3, \ldots$.

Ghasemi et al. [7] shows that the survival function of the assumed model is:

$$\begin{aligned} S(t, \theta(t)) = {} & \left[\prod_{l=0}^{k-1} \Pr\left(\theta^{l+1} | T > (l+1)\Delta, \pi^l\right)\right] \\ & \times \left[\prod_{l=0}^{k-1} \sum_{i=1}^{N} \pi_i^l \exp\left(-\int_{l\Delta}^{(l+1)\Delta} h(\tau, i)\, d\tau\right)\right] \\ & \times \sum_{i=1}^{N} \pi_i^k \exp\left(-\int_{k\Delta}^{t} h(\tau, i)\, d\tau\right) \end{aligned}$$

where:

$$\Pr\left(\theta | T > (l+1)\Delta, \pi^l\right) = \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_i^l p_{ij} q_{j\theta}, \tag{16.3}$$

and $h(\tau, i)$ is the PH model's hazard function at time τ while the equipment's degradation state is i. MLE is used to estimate the parameters of the model.

We consider a parametric PH-model with a baseline two parameter Weibull hazard function, which is also known as the Weibull parametric regression model (Banjevic et al. [8]), and is given as

$$h(t, Z_k; \beta, \eta, \gamma) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}\exp(\gamma Z_k); k\Delta \le t < (k+1)\Delta, k = 0, 1, \ldots. \quad (16.4)$$

16.3 Data Requirements for Parameter Estimations

Suitable data for the parameter estimation process of a prognostics model is derived from two sources:

- **Failure-related data:** This is based on the working age at which the equipment has failed. This information comes from installation and/or failure dates.
- **Performance data:** This is based on the condition indicators that relate to the performance condition of the equipment from its installation date up to its failure date obtained in fixed or near fixed observation intervals through condition monitoring systems.

In this research, the number of pass-by's is considered as the working age of a bearing. A pass-by is defined to be one pass in front of a specific condition monitoring station. By considering the pass-by as working age of the equipment, we address two issues in the modeling. First, the assumption of fixed or near fixed observation interval is satisfied. Second, short idle periods (when a wagon may be parked up or in the workshop) during a bearing's lifetime are not considered in their aging process.

16.3.1 Failure-Related Data

To develop the prognostics model, the following specific failure and historical data is necessary: (1) Exact Location of the bearing during its lifetime (wagon number, wheel set ID, left or right bearing on wheel set); (2) Installation date; (3) Removal date; (4) Failure or suspension; (5) Bearing Type (New, Over-hauled or Second-hand); (6) Failure codes.

There are >50,000 Rail car wheel bearing assemblies in service. Within this population there are three distinct bearing categories: New, Refurbished, and Second Hand. Bearings are recalled from service due to (1) hazardous condition monitoring values, or (2) a routine 2 year service. Part of this recall process is a hand-qualification and inspection test by the tradesman, following this, the bearing is sent to an external shop for refurbishing, or returned to service. The former is described here as 'refurbished' and the latter as 'second hand'.

In order to collect failure information, bearing failure has to be clearly defined. We have defined failure as follows. A bearing is failed if it was removed due to: (a) a high acoustic score (Roller, Cup, Cone); (b) a hot bearing detected alarm; (c) hand verification; (d) a catastrophic failure.

Failure data was collected on failures on a specific, single series of bearings over a 3 month period. This produced an initial 328 records of failure.

16.3.2 Modeling the Residual Life

Two most used measures of equipment health are the hazard function, and the MRL. These two measures are calculated from the reliability function.

In reliability analysis, two reliability functions are of interest. The first is the unconditional reliability function given by the probability $P(T > t)$, which is the probability that the failure time T, of a piece of equipment that has not yet been put into operation, is bigger than a certain time t. The second is the conditional reliability function calculated by $P(T > t|T > \tau)$, which is the probability that the time to failure T is bigger than t, knowing that the equipment has already survived until time τ, where $\tau < t$. In this later case, the MRL is $E(T - \tau|T > \tau)$ (Jardine et al. [9]), which is equal to $\int_{\tau}^{\infty} \frac{1-F(t)}{1-F(\tau)} dt, \tau > 0$, where T is having distribution function F. The hazard function $\lambda(\tau)$ is obtained from $\lambda(\tau) = P(\tau < T < \tau + \Delta\tau|T > \tau)/\Delta\tau$. In this equation $\Delta\tau$ is a small time interval less than the time to next inspection.

In some reliability analysis, it is assumed that every piece of equipment is used in the same environment, and under the same conditions. This assumption allows the calculation of the MRL, and the hazard function prior to the actual use of the equipment. In real-life, the environments in which the equipment is performing, and the conditions of utilization affect the process of degradation. Consequently, the conditional reliability, the residual life, and the failure rate of the equipment are affected. Taking this fact into consideration improves the diagnosis of the equipment's degradation state, and the prognosis for future performance.

Many researchers have proposed different reliability models incorporating the information gathered periodically regarding the equipment's condition. These models are used to calculate an adjusted hazard function, and the corresponding MRL. In this study, the PH-model with imperfect information is used. In the initial PH-model with perfect information, the conditional reliability is given by (Makis and Jardine [10]):

$$R(k, Z_k, t) = \exp\left(-\psi(Z_k) \int_{k\Delta}^{k\Delta+t} h_0(s)ds)\right), 0 < t \leq \Delta. \tag{16.5}$$

The conditional reliability indicates the probability of survival until time $k\Delta + t$, $(0 < t \leq \Delta)$, knowing that the failure has not happened until time $k\Delta$, and the states of the equipment have been $Z_1, Z_2, \ldots, Z_k$, at $\Delta, 2\Delta, \ldots k\Delta$. T is the random variable indicating the time to failure. Equation (16.5) is not valid for $t \geq \Delta$ because Z_k may change at any subsequent interval. In the case of imperfect information, Ghasemi et al. [11] has proved that the conditional reliability can be calculated by:

$$\bar{R}(k,\pi^k,t)=\begin{cases}\sum_{i=1}^{N}\pi_i^k\exp\left(-\psi(i)\int_{k\Delta}^{k\Delta+t}h_0(s)ds\right) & t\leq\Delta\\ \sum_{i=1}^{N}\pi_i^k R(k,i,\Delta)\sum_{j=1}^{N}p_{ij}\times R(k+1,j,(t-\Delta)) & t>\Delta\end{cases}. \tag{16.6}$$

And consequently, the MRL, $\bar{e}\left(k,\pi^k\right)$, calculated at the kth observation moment, while the state conditional probability is π^k, is calculated by:

$$\text{MRL} = \bar{e}\left(k,\pi^k\right)=\int_{k\Delta}^{\infty}\bar{R}\left(k,\pi^k,t\right)\,dt. \tag{16.7}$$

The steps for calculating the MRL at each observation moment k, where the indicator obtained is θ, are as follows. At any observation moment k, when an indicator value θ is obtained,

- Calculate the conditional probability distribution π^k, at period k using (16.1);
- Calculate the conditional reliability of the equipment, $\bar{R}(k,\pi^k,t)$, at the kth observation moment by using (16.7); and
- Calculate the MRL of the equipment, $\bar{e}\left(k,\pi^k\right)$, by applying.

16.4 Results

16.4.1 Performance Data Acquisition and Cleansing Outcomes

The original data set contained 328 failure records. However, in the cleansing process a large number of records had to be removed due to following reasons: (a) unknown age of the failed bearing, (b) unknown type of the bearing at installation moment (c) absence of historical acoustic data for long period, (d) inconsistency of dates in the databases

The outcome of the data acquisition and cleansing stage is a set of 63 records of failure, 57 for modeling purpose, and the 6 remaining records (almost 10 %) for testing the model result. These were removed from the data set and were not used to develop the model parameters. The distribution of bearing types for the 57 records is shown in Table 16.1.

Table 16.1 Average life of different bearing type

Bearing type	Average pass-by	STD-DEV	Count	% in Fleet (%)
New (N)	389	127	9	18
Overhauled (O)	291	95	5	23
Second-hand (S)	281	107	43	59
All	299	114	57	

Table 16.2 Model fitness with acoustic bearing indicators and bearing type

Parameter	Parameter estimation with LF and BT Estimate	Significant
Scale	610.10	–
Shape	2.96	–
State coefficient	0.43457[a]	–
Cup		N
Cone		N
Roller		N
Looseness		Y
Noisy		N
Bearing type		Y

[a] By considering just looseness and bearing type

16.4.2 Parameter Estimation Results

Table 16.2 shows the result of the parameter estimation method applied on collected data. Based on statistical tests, Cup; Cone; Roller: and Noisy energy bonds are not significant in the model. However, Looseness and Bearing Type are of great meaningful significant in the model. Bearing Type is almost ten times more significant than Looseness.

16.4.3 Testing a 15 Pass-by Replacement Policy

A replacement policy is criterion-based. For example, we decide whether to replace a bearing or leave it until the collection of next indicators. In this research we have introduced and tested following replacement policy: “Replace the bearing if x% lower confidence interval of MRL is less than 15 pass-bys” and found the best value of x that suits this replacement objective.

This criterion has been tested for confidence intervals of 68.3, 86.4, and 89.3 %. The results are summarized in Table 16.3. 68.3 % CI demonstrates 50 % True Negative (TN) and 50 % False Negative (FN) result. The FN ones are 31 pass-bys longer than real failure pass-by of the bearings. This means that in 50 % of cases the policy has estimated longer residual life for the bearings. On average the estimation has been 31 pass-by longer than real residual pass-bys.

By increasing the percentage of CI, at 89.3 % the results are satisfactory. Out of 12 decisions, 50 % have been correctly advised replacement (TP). 17 % of the cases did not need to be replaced and are correctly advised not to replace (TN). The remaining 33 % are the cases that the policy advises to replace the bearing before next 15 pass-by, while in reality the bearing could survive more than 15 pass-bys. However, the average difference of the estimation and the reality is 11

Table 16.3 15 Day alarm recall policy analysis for selected confidence intervals

Wheel	Policy: recall if X % lower CI is less than 15 Pass #	Real RL	X = 68.3	X = 86.4	X = 89.3
1	233	15	*Leave	*Leave	*Replace
	218	30	Leave	Leave	Replace
2	366	15	*Leave	*Replace	*Replace
	351	30	Leave	Replace	Replace
3	271	15	*Leave	*Leave	*Replace
	256	30	Leave	Leave	Leave
4	286	15	*Leave	*Leave	*Replace
	271	30	Leave	Leave	Replace
5	283	15	*Leave	*Replace	*Replace
	268	30	Leave	Replace	Replace
6	210	15	*Leave	*Leave	*Replace
	195	30	Leave	Leave	Leave
TP (* Replace)	0 %	17 %	50 %		
FP (Replace)	0 %	17 %	33 %		
TN (Leave)	50 %	33 %	17 %		
FN (* Leave)	50 %	33 %	0 %		
Average error (days)	31	−2	−11		

pass-by's. This means that the 33 % of the cases that have been advised to be replaced could survive in average 11 days more than predicted.

16.5 Discussion

In this part we explore the significance of the results in detail in each of the following five sub sections.

16.5.1 Outcomes of the Parameter Estimation

Table 16.2 shows the result of the parameter estimation method applied on collected data. Based on statistical tests, Cup; Cone; Roller; and Noisy energy bonds are not significant in the model. However, Looseness and Bearing Type are of great meaningful significant in the model. Insignificant behavior of these bearing acoustic indicators is contrary to expectations and may be due to one or combination of following reasons:

- Cone, Roller, and Cup were the indicators with the most missing data, i.e., values of zero in the field. These missing data were replaced with moving average;

- Looseness acoustic level of energy correlated with all or some of the four insignificant indicators, so it is reflecting their effect;
- There is much more significant information in Looseness indicator than imagined which is not being used currently;
- Insufficient number of failure data (samples);
- Error in data (this is discussed in the following section).

The significant of bearing type in the model is an important finding in the case and indicates that the maintenance staff should consider the consequences of selecting a second-hand bearing over a new or refurbished bearing.

16.5.2 Sample Population Characteristics

Table 16.1 summarizes the mean working age (Pass-by) of bearings grouped by different bearing types. The last column is an approximation of each bearing type percentage used in the fleet. This was calculated by using notification entries in the Maintenance Management System where the type of bearing installed each time is recorded.

Large standard deviation in the mean working age for two first types can be explained by small sample size, however second-hand bearings with 43 samples also show a large standard deviation value. The authors postulate that this is mostly due to diversity of second-hand bearings. In this case study, there is no record of the number of times a bearing has been used, remove and reinstalled as a second-hand bearing. A bearing classified as second-hand may have been reinstalled one to several times as a second-hand bearing. Also it may have been on other wheels as little as few weeks to several months. It also can be a second hand bearing of an originally new or overhauled bearing. Improvements in data collection, discussed in the recommendation section, should assist in accounting for this in future models.

16.5.3 Recommendations and Future Work

Recommendation and future works for this research work can be discussed in short and long terms. We suggest improving and applying the model by considering the following items:

- Extend data set: Currently, the model is constructed based on 57 failures and their related performance records. A first step to improve the output of the model is to extend the data to a larger set of failure data. At the same time better cleansing techniques and assumptions should be applied to the data. This will assure the integrity and correctness of the data fed to the model.

- Extend use of the acoustic energy levels available for modeling. In this work, we used 10 (5 for each bearing side) acoustic energy levels to feed the model. Ideally, one may consider using some/all 124 fields of acoustic signature produced by acoustic monitoring system to construct the model.
- Extend use of other degradation indicators such as Hot Bearing Detector and Wheel Impact Detector systems from the data acquisition, data cleansing, and modeling phases in this project.
- Develop data mining programs and AI tools in one package to automate the process of decision making.

16.6 Conclusions

We have presented a model for estimating the mean residual life of wagon bearings to support maintenance interventions. Data for the model has been collected from a rail industry case study. A prognostics model based on selected bearing acoustic data levels was developed. This provides an ability to detect failure in a 15 day window with acceptable levels of confidence.

Significant indicators that can be effectively used in calculation of MRL were distinguished and the MRL model was constructed based on these indicators. This case study has shown that (a) bearing type is of high significant impact in MRL and (b) some indicators like Looseness that have not been considered previously can carry very indicative information regarding the bearing MRL. We have also explored the frustrations and time consumption of data acquisition and data cleansing process and have suggested ways to address these problems. The positive impact of improving the data acquisition and cleansing process on MRL was discussed. Future work includes testing the significance of other condition monitoring indicators and also higher level of information from existing acoustic system data.

References

1. Wang W, Christer AH (2000) Towards a general condition based maintenance model for a stochastic dynamic system. Journal of the Operational Research Society 51:145–155
2. Makis V, Jiang X (2003) Optimal replacement under partial observations. Math Oper Res 28:382–394
3. Ghasemi A, Yacout S, Ouali MS (2007) Optimal condition based maintenance with imperfect information and the proportional hazards model. Int J Prod Res 45:989–1112
4. Ghasemi A, Yacout S et al (2010) Parameter estimation methods for condition based maintenance with indirect observations. IEEE Trans Reliab 59(2):426–439
5. Cox DR (1972) Regression models and life tables (with discussion). J R Stat Soc Ser B 26:187–220

6. Ghasemi A, Yacout S, Ouali MS (2008) Optimal strategies for non-costly and costly observations in condition based maintenance. IAENG Int J Appl Math 38(2):99–107
7. Ghasemi A, Hodkiewicz M, Salathiel G (2009) Report on Asset Selection for the Prognostics Project. UWA
8. Banjevic D, Jardine AKS, Makis V, Ennis M (2001) A control-limit policy and software for condition-based maintenance optimization. INFOR J 39:32–50
9. Jardine AKS, Banjevic D, Wiseman M, Buck S, Joseph T (2001) Optimizing a mine haul truck wheel motors' condition monitoring program. J Qual Maintenance Eng 7:286–301
10. Makis V, Jardine AKS (1992) Computation of optimal policies in replacement models. IMA J Math Appl Bus Ind 3:169–175
11. Ghasemi A, Yacout S et al (2010) Evaluating the reliability function and the mean residual life for equipment with unobservable states. IEEE Trans Reliab 59(1):45–54

Chapter 17
Detection of Failure in Gearbox Using Intensified Envelope Analysis

D. S. Gu, J. G. Kim, T. Kelimu, W. C. Kim and B. K. Choi

Abstract Acoustic Emission (AE) Method is widely used in research for machinery diagnostics, and wavelet transform has been implemented in many applications in the condition monitoring of machinery. In contrast to previous applications, this paper examines whether the acoustic signal can be used effectively to detect the fault in rolling element bearing using discrete wavelet transform (DWT). A novel mother function for DWT was extracted through acoustic signal measured by the fatigue crack growth test. A commonly encountered fault was simulated. The results suggest that DWT applied using the novel mother function is an effective for the detection of fault and may provide a powerful tool to indicate the faults in rolling element bearing.

D. S. Gu (✉) · J. G. Kim · T. Kelimu · W. C. Kim
Department of Mechanical and Precision Engineering, Graduate School, Gyeong-sang National University, Chundaegookchi-gil, Tongyeong City 650-160, Republic of Korea
e-mail: gdskam@gnu.ac.kr

J. G. Kim
e-mail: pjy1004@nate.com

T. Kelimu
e-mail: kream0401@gmail.com

W. C. Kim
e-mail: wckim@gnu.ac.kr

B. K. Choi
Department of Energy and Mechanical Engineering, Institute of Marine Industry, Gyeongsang National University, Chundaegookchi-gil, Tongyeong City 650-160, Republic of Korea
e-mail: bgchoi@gnu.ac.kr

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_17,

17.1 Introduction

Application of the high-frequency acoustic emission (AE) technique in condition monitoring of rotating machinery has been growing over recent years. This is particularly true for bearing defect diagnosis and seal rubbing [1–8]. The main drawback with the application of the AE technique is the attenuation of the signal and as such the AE sensor has to be close to its source.

Al-Ghamd and Mba [9] were focused on the application of the AE technique for identifying the presence and size of a defect on a radially loaded bearing. They concluded that the fundamental source of AE in seeded defect test was due to material protrusions above the mean surface rough-ness; AE r.m.s. maximum amplitude and kurtosis have all been shown to be more sensitive to the onset and growth of defects than vibration measurements. In addition, a relationship between the AE burst duration and the defect length has been presented. In a separate study, Elforjani and Mba [10] demonstrate the use of AE measurements to detect, monitor the growth, and locate natural defect initiation and propagation in a conventional rolling element bearing. In the research, it has been shown that the AE energy is a reliable, robust, and sensitive parameter for detection of incipient cracks and surface spalls in a slow speed bearing while in operation. And it also successfully demonstrated the ability to determine the source of AE during operation. In an additional study by Al-Balushi et al. [11], he presents results on the Energy Index (EI) technique, used in detecting masked AE signatures associated with the loss of mechanical integrity in bearings, and concluded that the EI technique is effective in detecting AE burst buried in random noise, thereby offering a complementary tool for the diagnostician. Other research has shown the ability and the effectiveness of AE technique using acoustic parameters trough experimental method. Kim et al. [12] presented the fault detection of the rolling element bearing through envelope analysis and wavelet transform. It was concluded that defect frequencies can be detected in the power spectrum the peak of which was dominated by wavelet filter. Gu et al. [13–15] applied the wavelet filter to identify the characteristics of the crack growth. In the result, he also showed that wavelet filter for envelope analysis can be used for a high quality de-noising technique.

Therefore, in this paper, a novel scaling filter for DWT (Discrete Wavelet Transform) will be proposed through fatigue crack growth test using a specimen made from SM53C and applied to detect the faults in a gearbox. To define the defect frequencies, envelope analysis was used for signal processing. For the test, the gearbox was operated for 15 days with misalignment generated between a gear and pinion, and AE signals were acquired with 5 MHz sampling rate for every 30 min.

17.2 Development of a Novel Mother Function

17.2.1 Fatigue Crack Growth Test

The material which is used in this experiment with the mechanical structural carbon steel (SM53C) is widely used, and SM53C was head of high-class carbon steel which is provided in KS standard D3752. Figure 17.1 shows the form and size of the specimen for fatigue crack growth test. The tensile period was 10 Hz sinusoidal with 3 mm and 0.1 stress ratio (R).

An AE sensor of wideband type attached on the side of the specimen, pre-amplifier, DAQ board, and signal analysis system of PAC (Physical Acoustic Corporation) were used. The detail system information and the setting value of system parameters are shown in Table 17.1. The AE signal was stored 0.4 milli-seconds (ms) by time-based trigger with 30 s until separation of the specimen, and it has taken time for about 100,000 cycles to sever.

17.2.2 Test Results

Figure 17.2 shows a time wave form caused by crack growth, which was extracted by time interval from the maximum peak value. It was shown that a high frequency signal was riding on a low frequency signal (two dash lines on Fig. 17.2) from the occurrence of a dominate peak. This signal was observed during the whole test, its observing period was increased along the crack growth. Johnson [16] also observed the kind of signal through crack growth test using composite plate as shown in Fig. 17.3.

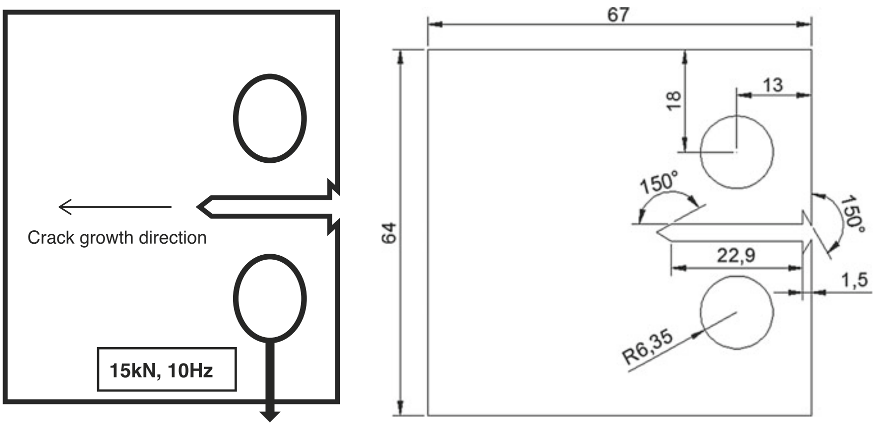


Fig. 17.1 Load condition and design of the specimen

Table 17.1 Specification of the AE sensor

AE sensor (wideband type)	Peak sensitivity V/(m/s); [V/μbar]: 55 [−62] dB Operating frequency range: 100–1,000 kHz Resonant freq. V/(m/s); [V/μbar]: 125 [650] kHz Directionality: ±1.5 dB

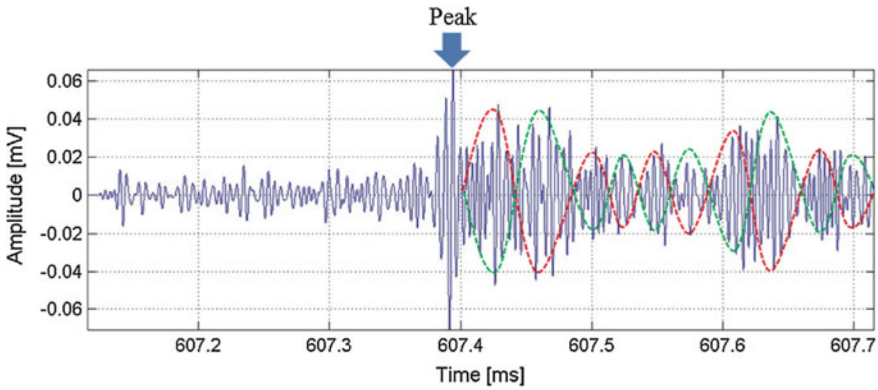


Fig. 17.2 Time waveform of the AE signal caused by the crack growth

Figures 17.2 and 17.3 exhibit a similar trend as follows:

- Occurrence of a high peak at the start an increase in amplitude.
- Creation of a high frequency signal just prior to the generated peak.
- Consistent change in signal over a long period of time following the generated peak.

A novel mother function for AE signal was developed as shown in Fig. 17.4 with three conditions: summation of the pattern equals 1, root mean square (RMS) value is 0.7 and length is 2N (N = constant) [17].

To validate the proposed pattern, it was applied for crack growth test signals. Figure 17.5 shows the time waveform and power spectrum of the same signal as shown in Fig. 17.2. In the power spectrum of Fig. 17.5, we can find some peaks that were 15.5, 29.8, 45, 59.6, and 75.1 Hz. The first peak is the operating period dominated in the time waveform and the others are harmonics components of operating frequency. In general, rotating machine harmonics caused by faults are observed in critical components of the machine such as bearing, rotor of electric motor and shaft. In this test, the specimen vibrated by force system and the operating signal can be demodulated by the crack growth as it is regarded as a kind of fault in the test system. So, the harmonics of the operating frequency occurred. In addition, 29.8 Hz (2X of the operating speed) is bigger than the first peak because it could consider that exiting point, bottom and top position of the forcing bar. The wavelet filter applied to the suggested pattern is working for signal processing of the AE signal.

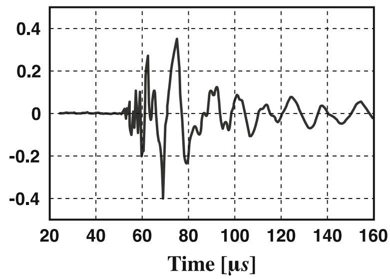


Fig. 17.3 AE recording [16]

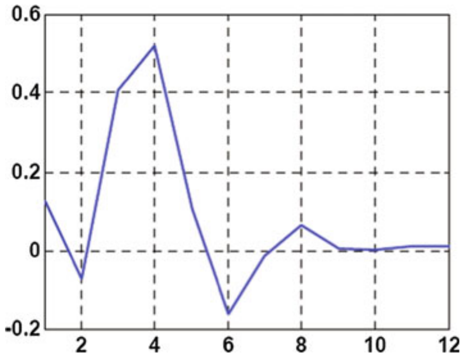


Fig. 17.4 Proposed mother function

17.3 Intensified Envelope Analysis

Figure 17.6 shows a flow chart of Intensified Envelope Analysis (IEA). General enveloping method consists of band-pass filter (BPF), rectification, Hilbert transform, low-pass filter (LPF), and FFT. A critical problem of general enveloping method is in the BPF because the cutoff frequencies of the BPF are decided without any references and/or standards and the results of the enveloping is seriously affected by the cutoff frequencies. In addition, it is widely known that it makes a best result when the cutoff frequencies include the natural frequency of the system.

Therefore, in this paper, DWT is used to remove the disadvantage of envelope analysis, and a novel mother function is applied to increase the filtering effect focused on AE signal. DWT is worked as filter and denoising by the proposed mother function. The AE signal which was preprocessed by IEA has lower level of the background noise and stronger fault signal than the result of the original enveloping method.

17.4 Application of the Novel Mother Function

The test-rig employed for this investigation consists of one identical oil-bath lubricated gearbox, 3 HP-motor, rigid coupling, tapper-roller bearing, pinion, gear, control panel, and break system, as seen Fig. 17.7. The pinion was made from steel

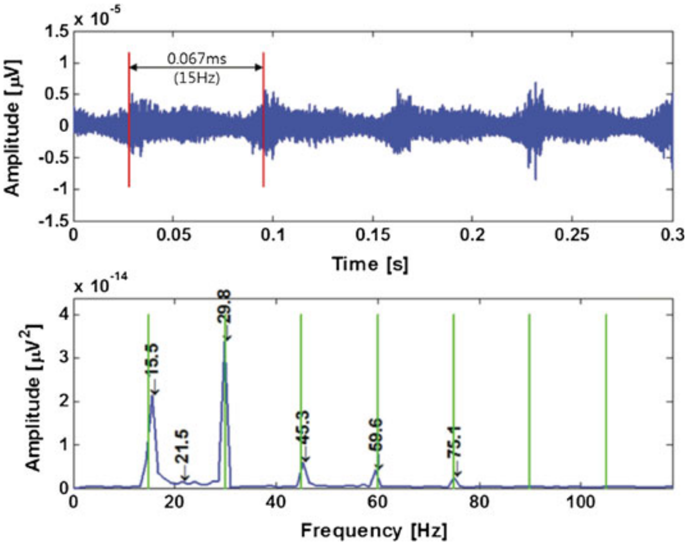


Fig. 17.5 Time waveform and power spectrum of the crack growth test signal

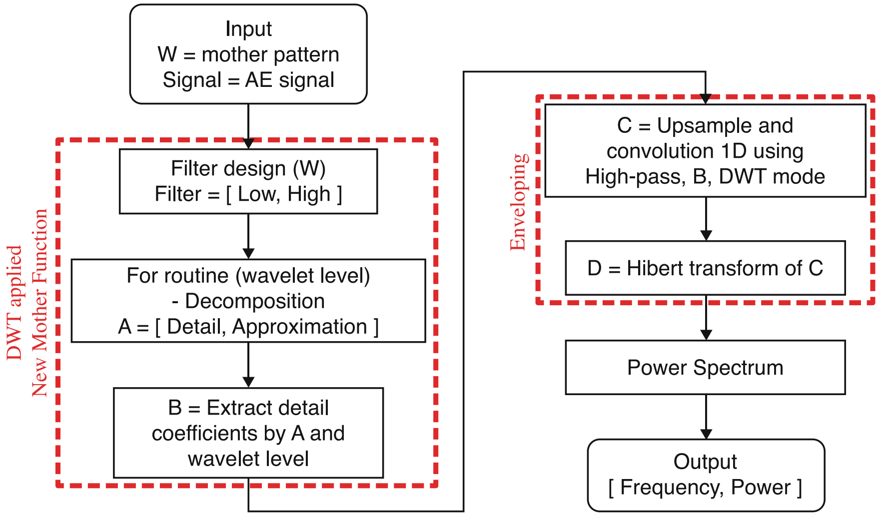


Fig. 17.6 Flow chart of IEA

with heat treatment, the number of teeth is 70, and diameter is 140 mm. The gear was made from steel, but it was produced without any heat treatment process during manufacturing. The number of gear teeth is 50 and diameter is 100 mm, and module is 2 mm for the gear and pinion, respectively.

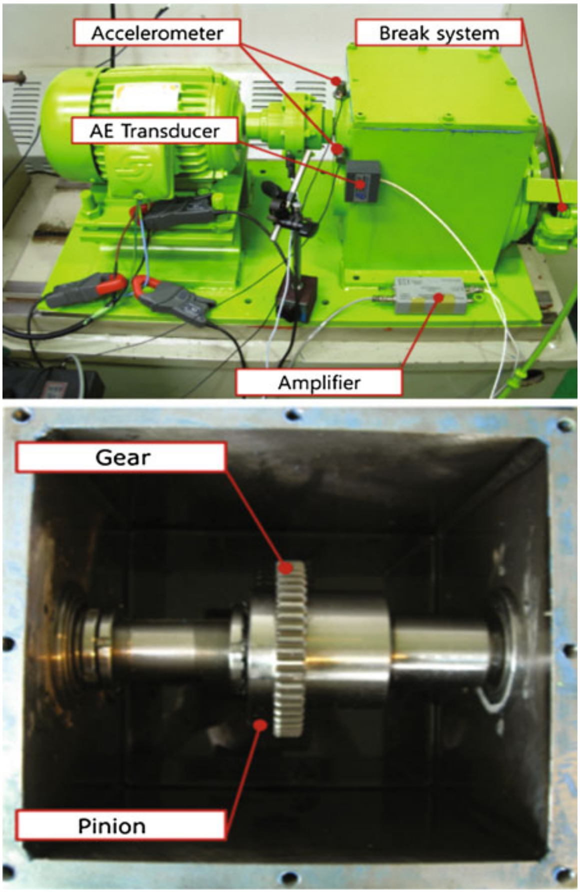


Fig. 17.7 Test rig

Table 17.2 Specifications of the gearbox

	Gear	Pinion		
No. of teeth	50	70		
Speed of shaft	25.01 rev/s			
Meshing frequency	1,250 Hz	1,750 Hz		
		Bearing (NSK HR 32,206 J)		
No. of rolling element	17	Type	Defect Freq.	Fault Freq.
Diameter of outer race	62 mm	BPFO	8.76 Hz	219.3 Hz
Diameter of inner race	30 mm	BPFI	11.24 Hz	281.38 Hz
		FTF	3.84 Hz	96.13 Hz
		BSF	0.44 Hz	11.01 Hz

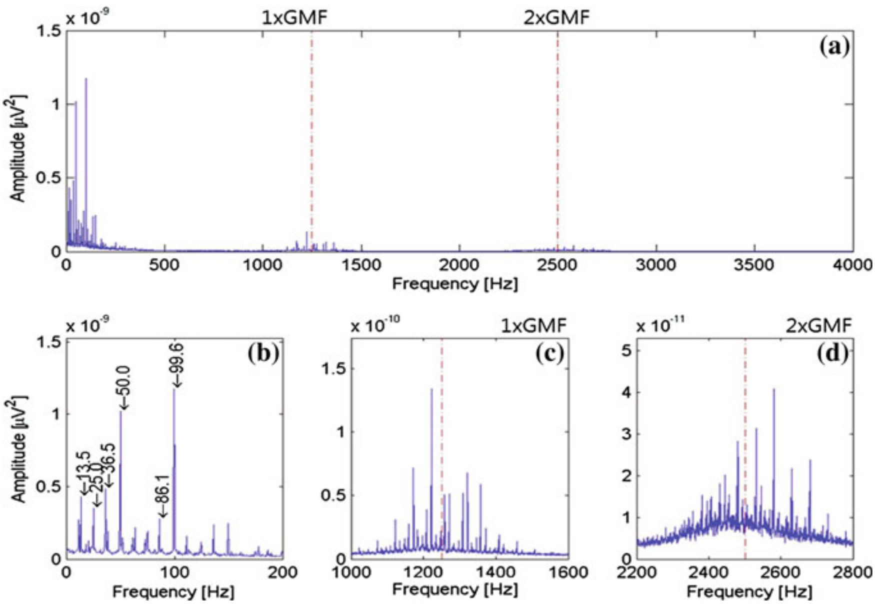


Fig. 17.8 Averaged power spectrum and zooming of 14th day

A simple mechanism that permitted a break of disk-pad type to be rotated relative to each other was employed to apply torque to the gear. Contact ratio (Pinion/Gear) of the gears was 1.4. The motor used to drive the gearbox was a 3-phase induction motor with a maximum running speed of 1,800 rpm, respectively, and was operated for 15 days with 1,500 rpm. The torque on the output shaft was 1.2 kN·m while the motor was in operation, and other specifications of the gearbox are given as in Table 17.2. [17, 18]

In the results of the envelop analysis with DWT, the high harmonics of f_m occurred by strong wearing phenomena caused by misaligned teeth. In the power spectrum as shown in Fig. 17.8. 25 Hz (f_r) and its harmonics are generated and 11.32 Hz was the ball pass frequency of inner race (BPFI [f_d]). In Fig. 17.8c and d, the center dash line is shown for f_m and $2f_m$, and their side lines are the sidebands with difference 25 Hz (f_r).

17.5 Conclusions

Defect simulation for critical components of general machinery had been performed and analyzed using IEA, which is envelope analysis intensified by DWT with a novel mother function. The result of each case is shown as following:

- The characteristic frequencies of defects and their harmonics were clearly indicated on every case.
- Misalignment was observed and bearing faults were also detected in the early fault stage.
- IEA is utilized to evaluate the faults and indicate the fault frequencies, rotating speed, sideband, gear mesh frequency, and harmonics in the gearbox test, explicitly.

According to the results, IEA can be a powerful signal processing method for condition monitoring to perform early fault detection. Most notably, it has a better weighting function for harmonics and sidebands of defect characteristic frequencies than enveloping as shown in the test results. That can be connected with the early fault detection because it can represent the defect frequencies of initial faults. In addition, the novel mother function is optimized and focused on some defects of rolling element bearing and gear.

Acknowledgments This research was supported by Basic Science Research Program through the National Re-search Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0013652) and the 2nd Phase of Brain Korea 21.

References

1. Mba D, Bannister RH, Findlay GE (1999) Condition monitoring of low-speed rotating machinery using stress waves: Part 1 and Part 2. Proc Inst Mech Eng, Part E: J Process Mech Eng 213(3):153–185
2. Holroyd T, Randall N (1993) The use of acoustic emission for machine condition monitoring. Braz J Non-Destruct Test 35(2):75–78
3. Mba D (2001) The detection of shaft-seal rubbing in large-scale turbines using acoustic emission. 14th international congress on condition monitoring and diagnostic engineering management. Manchester. pp 21–28
4. Mba D, Cooke A, Roby D, Hewitt G (2003) Opportunities offered by acoustic emission for shaft-steal rubbing in power generation turbines: a case study. International conference on condition monitoring, Queensland. pp 280–286
5. Kim YH, Tan ACC, Mathew J, Kosse V, Yang BS (2007) A comparative study on the application of acoustic emission technique and acceleration measurements for low speed condition monitoring. 12th Asia-Pacific vibration conference. Sapporo
6. Kim YH, Tan ACC, Mathew J, Yang BS (2007) Experimental study on incipient fault detection of low speed rolling element bearings: time domain statistical parameters. 12th Asia-Pacific vibration conference. Sapporo
7. Tan CK, Mba D (2005) Limitation of acoustic emission for identifying seeded defects in gearboxes. J Non-Destructive Eval 24:11–28
8. Siores E, Negro AA (1997) Condition monitoring of a gear box using acoustic emission testing. Mater Eval 55:183–187
9. Al-Ghamd AM, Mba D (2006) A comparative experimental study on the use of acoustic emission and vibration analysis for bearing identification and estimation of defect size. Mech Syst Signal Process 20:1537–1571
10. Elforjani M, Mba D (2009) Detecting natural crack initiation and growth in slow speed shafts with the Acoustic Emission technology. Eng Fail Anal 16:2121–2129

11. Al-Balushi KR, Addali L, Charnley B, Mba D (2010) Energy index technique for detection of Acoustic Emission associate with incipient bearing failures. Appl Acoust 71:812–821
12. Kim BS, Gu DS, Kim JG, Kim YC, Choi BK (2009) Rolling element bearing fault detection using acoustic emission signal analyzed by envelope analysis with discrete wavelet transform. Eng Asset Lifecycle Manage 2010:596–602
13. Gu DS, Kim BS, Kim JG, Choi BK (2009) Low speed shaft fault detection using acoustic emission. The 13th Asia-pacific conference on non-destructive testing. Yokohama
14. Gu DS, Kim JG, Park WJ, Choi BK (2010) Cracked shaft condition monitoring using acoustic emission system. The 17th international congress on sound and vibration. Cairo
15. Gu DS, Kim JG, Kim YC, Park WJ, Choi BK (2010) Feature estimation of AE signals caused by rotating shaft with crack growth. The 6th international conference on physical and numerical simulation of materials processing. Guilin
16. Johnson M (2003) Classification of AE transient based on numerical simulations of composite laminates. NDT&E Int 36:319–329
17. Gu DS, Kim JG, An YS, Choi BK (2011) Detection of faults in gearboxes using acoustic emission signal. J Mech Sci Technol 25(5):1279–1286
18. Lee JH, Gu DS, Jang YS, Choi BK, Kim WC, Son JD, Yang BS, Widodo A (2008) Detection of the gear fault via acoustic emission using features and wavelet transform. Proceedings of the 3rd World congress on engineering asset management and intelligent maintenance systems. Beijing, pp 846–852

Chapter 18
The AE Signal Analysis of the Fatigue Growth Test Using Acoustic Emission

J. G. Kim, D. S. Gu, H. Hwang, W. C. Kim and B. K. Choi

Abstract The acoustic emission (AE) technique is a well-known non-destructive test which is widely applied in research and industry. It is mainly used in monitoring the onset of the cracking process in materials to predict and prevent the factures that attract the attention of researchers to apply condition monitoring techniques into detecting the early stage of cracks. The object of this study is to obtain the features of AE signals caused by crack growth. The envelope analysis with discrete wavelet transform (DWT) is used to find the characteristic of AE signal. To estimate the features, the crack length was divided into three parts. From the raw signals and transferred signals, the power spectrum which was processed by the envelope analysis with DWT was generated. Through this experiment, the envelope analysis with DWT is a better way to analyze the original AE signals. In this paper, the results of fatigue crack growth can contribute to a database of crack detection.

J. G. Kim · D. S. Gu · H. Hwang
Department of Precision Mechanical Engineering, Gyeongsang National University,
445, Inpyeong-dong, Tongyeong, Gyeongsangnam-do 650-160, South Korea
e-mail: pjy1004@nate.com

D. S. Gu
e-mail: gdskam@gnu.ac.kr

W. C. Kim
Department of Mechanical System Engineering, Gyeongsang National University,
445, Inpyeong-dong, Tongyeong, Gyeongsangnam-do 650-160, South Korea
e-mail: wckim@gnu.ac.kr

B. K. Choi (✉)
Department of Energy and Mechanical Engineering, Institute of Marine Industry,
Gyeongsang National University, 445, Inpyeong-dong, Tongyeong,
Gyeong-sangnam-do 650-160, South Korea
e-mail: bgchoi@gnu.ac.kr

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_18,

18.1 Introduction

The acoustic emission (AE) technique is a well-known non-destructive test which are widely applied in research and industry. It is mainly used in monitoring the onset of the cracking process in materials to predict and prevent the factures that attract the attention of researchers to apply condition monitoring techniques into detecting the early stage of cracks. One of AE's great advantages is the ability to scan a large structure with only a few sensors and localize the source caused by the captured AE transient. AE is a term used to describe a class of phenomenon in which transient elastic waves are generated by the rapid release of energy from localized sources within or on the surface of a material. So far, by using the AE technique, fatigue crack research had focused on the relationship between the stress intensity factor and crack length. However, fatigue crack research methods have been evolving. signal processing studies are underway that have applied vibration signal analysis methods (power spectrum, envelop analysis, wavelet transform, etc.) at acquired signals [1–3].

The objective of this research is to obtain the features of AE signal caused by crack growth. To understand waveforms of the AE signal, fatigue crack growth test has been performed using the specimen made from general steel (SM45C). From the raw signals and transferred signals, the power spectrum which was processed by the envelope analysis with DWT was generated. In the result of this research, the cracked specimen was fractured around 150 min within operating time. The crack length was acquired by using an extensometer, and it was divided into three parts by crack length to estimate the features. This research will contribute to the data for crack detecting.

18.2 Experimental Methods

The specimen was made from SM45C satisfied KS standard D 3752. Experiment condition using servo hydraulic test system was 15kN of load and 0.1 of stress ratio (R) with 10 Hz exciting frequency. AE systems (AE converter, amplifier, DAQ board, and analysis software) are the products of Physical Acoustic Corporation (PAC) company. The amplifier has a scope in 20 kHz–1.2 MHz, and set with 40 dB. The AE transducer was attached by using the magnetic holder on the end of the crack growth direction. The AE signals were stored by sampling frequency of 5 MHz at every 10 s with a time based trigger. An extensometer (Clip-on Displacement Gauge) was used for measuring the crack length. Table 18.1 presents the experiment AE system parameters and Table 18.2 is the specification of the data acquisition system (Figs. 18.1, 18.2).

Table 18.1 Parameters setting of AE

Parameter type	Hardware setup value
Threshold value	30 dB
Preamplifier	40 dB
Peak definition time (PDT)	300 μs
Hit definition time (HDT)	600 μs
Hit lockout time (HLT)	1,000 μs
Sample rate	5MSPS
Pre-trigger	100 μs
Hit length	40 μs
Filter on board (Low)	1 kHz
Filter on board (High)	2 MHz

Table 18.2 Specification of data an acquisition system

2 channel AE system on PCI-board	18-bit A/D conversion 10 M samples/sec rate (on one channel, 5 M samples/sec on 2 AE channels)
AE sensor (wideband type)	Peak sensitivity V/(m/s); [V/μbar]…55[−62] dB
	Operation frequency range…100–1,000 kHz
	Resonant frequency V/(m/s); [V/μbar]…125[650] kHz
	Directionality…±1.5 dB
Preamplifier (20/40/60 dB gain)	Wide dynamic range < 90 dB
	Signal BNC 20/40/60 selectable gain

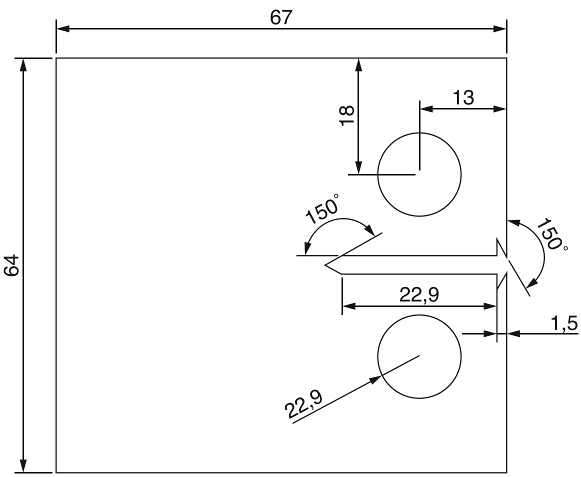


Fig. 18.1 Experimental model

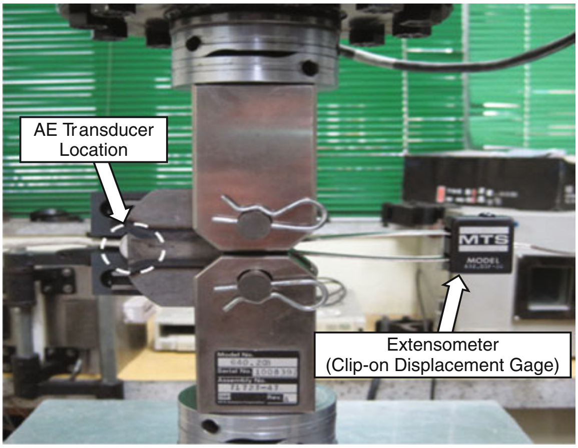


Fig. 18.2 Transducer location on the test specimen

18.3 Signal Processing

18.3.1 Envelope Analysis

Envelope analysis is used in detecting fault signal periodically, which is mixed with noise signals. This method has the ability to separate the background noise signals in the AE signals. The sequences of a typical envelope analysis procedure are: (1) band-pass filter, (2) wave rectification, (3) Hilbert transform or low-pass filter, (4) power spectrum. The purpose of band-pass filtering is to reject the low-frequency high-amplitude signals associated with mechanical vibration components and to eliminate random noise outside the pass-band. So, it's very important to set up bandwidth in envelope analysis [4].

Range of BPF is performed in five ranges (10–50 kHz, 50–100 kHz, 100–300 kHz, 300–500 kHz, and 500–800 kHz), and among them the best signal characteristics appears at the range of 100–300 kHz.

18.3.2 Discrete Wavelet Transform

Wavelet analysis, also called wavelet transform, is a remarkably powerful and general method for its many distinct merits. Newland's work [5] made it popular in engineering applications, especially for vibration analysis [6–9].

There are two kinds of wavelet transform: continuous wavelet transform (CWT) and discrete wavelet transform (DWT). CWT is defined as the sum over all time of the signal multiplied by scaled, shifted versions of the wavelet function. To use

CWT, one signal can be decomposed into a series of "small" waves belonging to a wavelet family composed of scaling functions.

For many signals, the low-frequency content is the most important component, as it gives the signal its identity. The high-frequency content, on the other hand, imparts the flavor or nuance that is often useful for singular signal detection. In wavelet analysis, we often speak of approximations and details. The approximations are the high-scale, low frequency components of the signal. The details are the low-scale, high frequency components.

However, CWT is time intensive due to calculating the wavelet coefficient at all scales and it produces a lot of data. To overcome such a disadvantage, we can choose scales and positions based on powers of two so-called dyadic scales and positions—when wavelet analysis will be much more efficient and just as accurate. Such an analysis is obtained from the discrete wavelet transform (DWT). The approximate coefficients and detail coefficients decomposed from a discrete signal can be expressed as:

$$a_{(j+1),k} = \sum\nolimits_{k=0}^{N} a_{j,k} \int \phi_{j,k}(t) \cdot \phi_{(j+1),k}(t)\mathrm{d}t = \sum\nolimits_{k} a_{j,k} \cdot g[k]. \tag{18.1}$$

$$d_{(j+1),k} = \sum\nolimits_{k=0}^{N} a_{j,k} \int \phi_{j,k}(t) \cdot \psi_{(j+1),k}(t)\mathrm{d}t = \sum\nolimits_{k} a_{j,k} \cdot h[k]. \tag{18.2}$$

The decomposition coefficients can therefore be determined through convolution and implemented by using a filter. The filter, $g[k]$, is a low-pass filter and $h[k]$ is a high-pass filter. The decomposition process can be iterated, with successive approximations being decomposed in turn, so that one signal is broken down into many lower resolution components. This is called the wavelet decomposition tree as shown in Fig. 18.3.

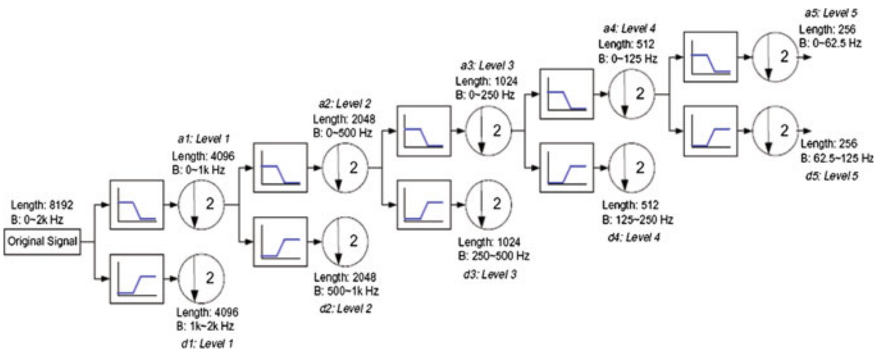


Fig. 18.3 Wavelet decomposition tree

18.4 Experiment Results

The specimen was fractured around 150 min into its operating time. When the crack is growing, it is happened in inside faster than surface of the specimen [10]. Figure 18.4 shows the crack length along the number of cycle. To consider the feature of the AE signal along crack growth, crack growth was divided into three parts by crack length.

Region 1. crack length 0–1.5 mm (0–33,800 cycle)
Region 2. crack length 1.5–9.5 mm (33,800–83,895 cycle)
Region 3. crack length 9.5–14.5 mm (83,895–90,225 cycle)

To estimate the feature of each region, the time waveform and the power spectrum were analyzed and compared. Figure 18.5 shows the time waveform and the power spectrum of each region processed by envelope analysis only. At the region 1, the amplitudes were appeared in 110–130 kHz, 150–170 kHz, 190–210 kHz, and 250–270 kHz. At the region 2, 160 kHz's amplitude was increased rapidly. At region 3, all amplitudes that appeared in region 1 were increased but 160 kHz which is decreased. Envelope analysis with DWT is used to convert a known frequency range in machinery condition monitoring because it is not known that the mechanism will bring out these frequencies.

Figure 18.6 is the power spectrum of the AE signal processed by envelope analysis with DWT. X-axis is frequency (Hz) and Y-axis is amplitude (mV^2). In Fig. 18.6a, 9.54 Hz component is appeared lower than 19.07 Hz component. In Fig. 18.6b, the 9.54 Hz component was increased and it was appeared higher than 19.07 Hz component. In Fig. 18.6, 9.54 Hz component was decreased and harmonic components of 10 Hz was appeared. In Fig. 18.6, 30.99, and 40.53 Hz component was increased, and the peaks of 50 Hz interval from the 100 Hz were appeared to decrease from 107 to 121 min.

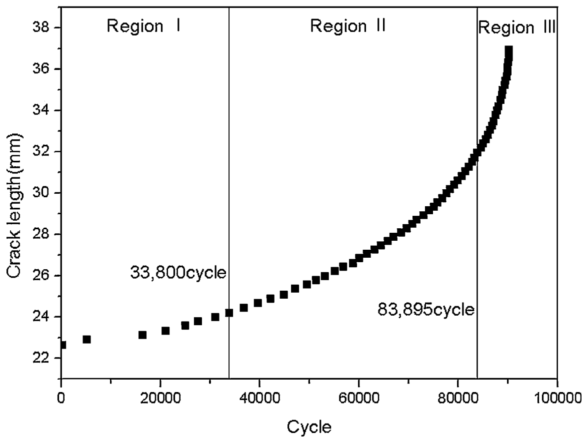


Fig. 18.4 Correlation of the cycle and the crack length

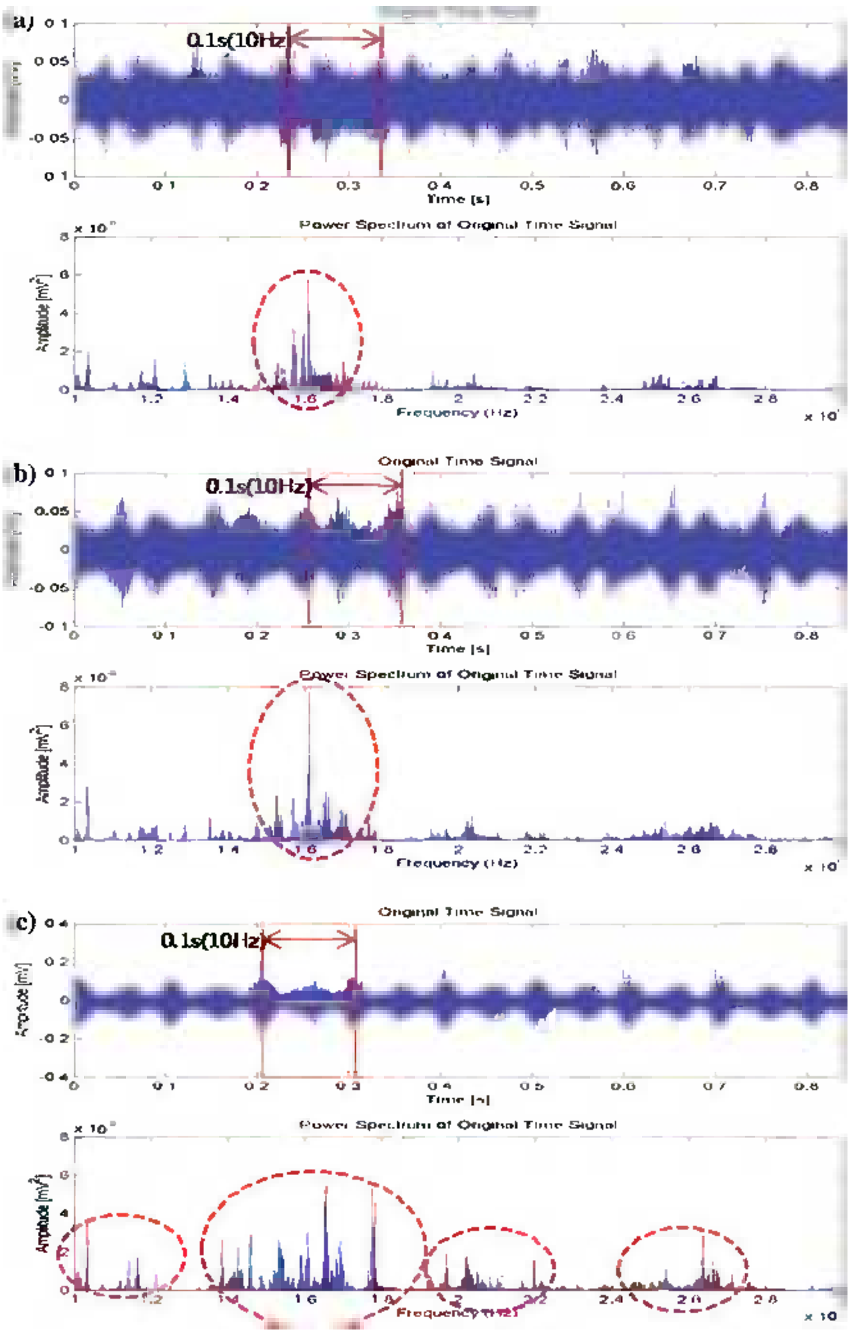


Fig. 18.5 Time waveform and power spectrum of the AE signal envelope analysis without DWT (100–300 kHz) (**a** Region 1, **b** Region 2, **c** Region 3)

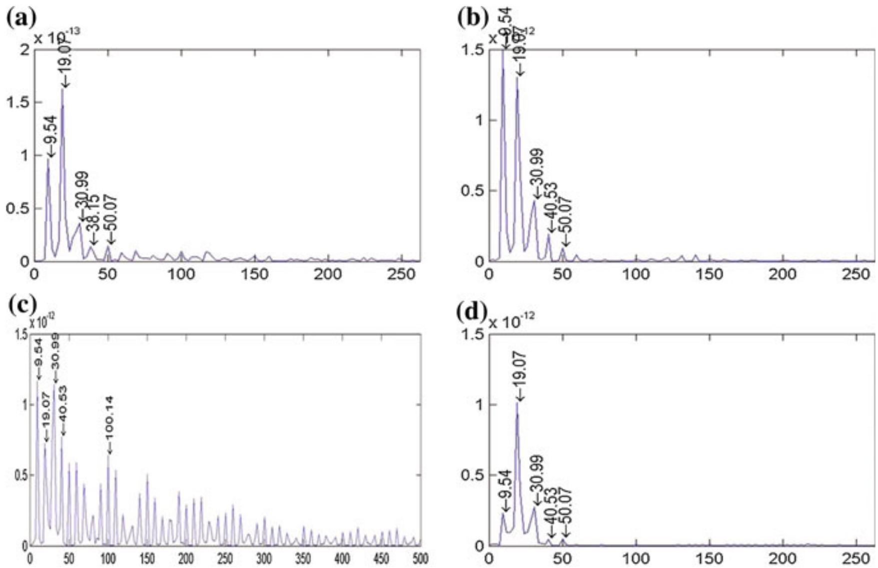


Fig. 18.6 Power spectrum of the AE signal envelope analysis with DWT. [**a** Region 1, **b** Region 2, **c** Another feature of 14 min in region 2 (64,200–72,600 cycle), **d** Region 3]

In signal processing, STFT is the process of reducing the sampling rate of the signal. It is usually done to reduce the rate or the size of the data. Figure 18.7 is STFT of the AE signal envelope analysis with DWT. Each region follows the pattern as can be seen in Fig. 18.7, and the last 14 min of the pattern in region 2 moves to region 3. The frequency shows around 160 kHz at region 1, and region 2 shows 160 kHz when the center frequency of region 2 is 200 kHz. Region 3 is similar to region 2, except that there is another 200 kHz frequency located in front of the center frequency. Figure 18.7c follows a different pattern wherein only center frequencies exist. And like Figs. 18.6c, 18.7c exhibits different characteristics in power spectrum of the AE signal by envelope analysis without DWT.

The characteristics of the AE signals are difficult to identify due to the mixing high frequency noise signals in the spectrum without DWT. However, the different features acquired during the last 14 min at the region 2 is the processing of the power spectrum and the STFT of the AE signal by envelope analysis with DWT.

18.5 Conclusion

In this paper, the object of the study is to obtain the features of AE signal caused by crack growth. To understand waveforms of the AE signal, fatigue crack growth test has been performed using the specimen made from general steel (SM45C).

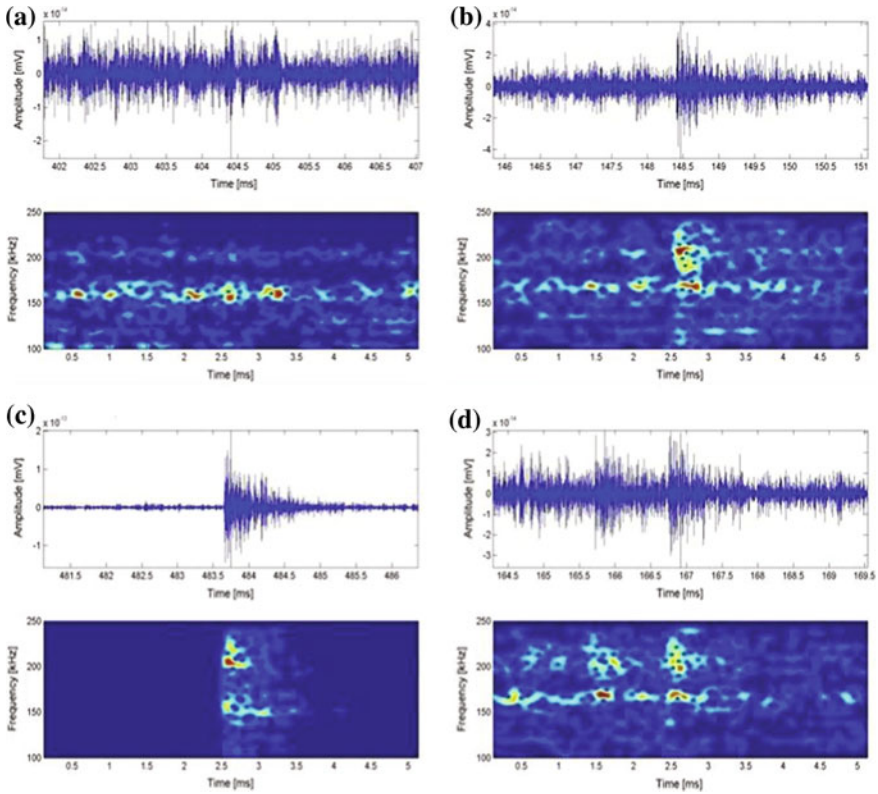


Fig. 18.7 STFT of the AE signal envelope analysis with DWT [**a** Region 1, **b** Region 2, **c** Another feature of 14 min in region 2(107–121 min), **d** Region 3]

AE sensor is used in capturing the signal in the fatigue test. The time waveform and the power spectrum were obtained from the AE signals, and characteristics of the signal are analyzed. However, the characteristics of the signal were difficult to analyze due to the mixing high frequency noise signal in the power spectrum without DWT. After applying the envelope analysis with DWT method, the noise signal had been removed and the analysis of the characteristics of AE signals was possible. Through this experiment, the envelope analysis with DWT is a better way to analyze the original AE signals. In this paper, the results of fatigue crack growth can contribute to database of crack detection.

Acknowledgments This research was financially supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) founded by the Ministry of Education, Science, and Technology(2011-0013652), and was supported by The Eco-Friendly Heat & Cold Energy Mechanical Research Team of the 2nd Phase of Brain Korea 21.

References

1. Johnson M, Gudmundson P (2000) Broad-band transient recording and characterization of acoustic emission events in composite laminates. Compos Sci Technol 60(2):2803–2818
2. Johnson M (2003) Classification of AE transients based on numerical simulations of composite laminates. NDT&E Int 36:319–329
3. Åberg M (2001) Numerical modeling of acoustic emission in laminated tensile test specimens. Int J Solids Struct 38:6643–6663
4. Jun OS (2008) Cracked rotors: Industrial machine case histories and nonlinear effects shown by simple Jeffcott rotor. Mech Sys Sig Process 22:805–817
5. Cempel C (1991) Vibroacoustic condition monitoring. Ellis Horwood, Chichester
6. Tang KS, Man KF, Kwong S, He Q (1996) Genetic algorithms and their applications. IEEE Sig Process Mag 13(6):22–37
7. Jack LB, Nandi AK (2000) Genetic algorithms for feature selection in machine condition monitoring with vibration signals. IEEE Proc Vis Image Sig Process 147(3):205–212
8. Jack LB, Nandi AK (2002) Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms. Mech Sys Sig Process 16:373–390
9. Samanta B (2004–2009) Artificial neural networks and genetic algorithms for gear fault detection. Mechanical systems and signal processing. Mech Sys Sig Process 18:1273–1282
10. Nam KW, Kim SJ (1998) "Signal characteristics of acoustic emission from fatigue crack propagation", KSME, Proceedings of the KSME annual autumn conference, pp. 53–58

Chapter 19
Commercialization of Prognostics Systems Using Commercial Off-the-Shelf Technologies

P. Johnson

Abstract There are many facets of a prognostics and health management system. Facets include data collection systems that monitor machine parameters; signal processing facilities that sort, analyze, and extract features from collected data; pattern matching algorithms that work to identify machine degradation modes; database systems that organize, trend, compare, and report information; communications that synchronize prognostic system information with business functions including plant operations; and finally visualization features that allow interested personnel the ability to view data, reports, and information from within the intranet or across the Internet. A prognostic system includes all of these facets, with details of each varying to match specific needs of specific machinery. To profitably commercialize a prognostic system, a generic yet flexible framework is adopted which allows customization of individual facets. Customization of one facet does not materially impact another. This chapter describes the framework, customization process, and choices of commercial system components.

19.1 Introduction

The objective of a prognostic system is to predict the need for maintenance before serious equipment breakdown occurs and to predict the remaining useful life of the equipment components. A prognostics system should operate to some extent on its own, to lessen the need for human involvement. In order for a prognostic system to work somewhat autonomously, each component should work together seamlessly.

Prognostics Systems have several components, commonly grouped into data acquisition, signal processing, and analysis, and decision making, Fig. 19.1.

P. Johnson (✉)
National Instruments Corporation, 11500 North Mopac Blvd, Austin, TX 78759, USA
e-mail: preston.johnson@ni.com

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_19,

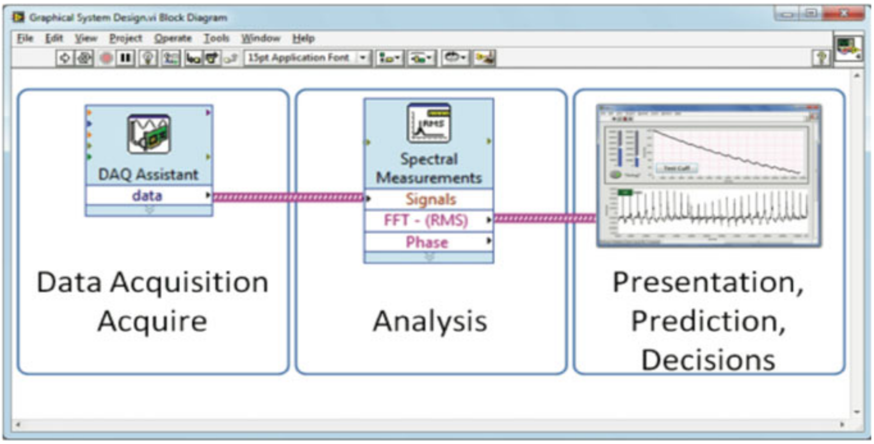


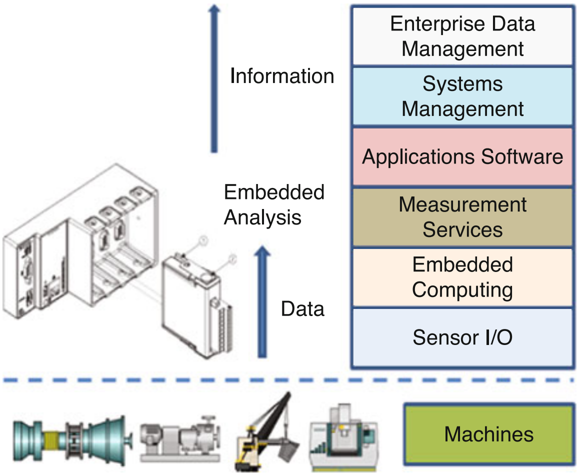
Fig. 19.1 Components of prognostic system

Figure 19.1 can be expanded to include communications, visualization, and database components.

To allow flexibility for analysis types, and machinery types, each component needs to be modular to the extent components can be easily interchanged. This interchangeability extends to hardware as well as software. The system needs to scale from small machines up to large machines, and from test cell applications down to portable systems and into online embedded systems.

Any online embedded system components need to be priced competitive to existing data collecting systems. In other words, constraints exist in hardware, software, and development tools in order to maximize modularity, cost, and ease of customization. A framework of hardware and software components makes this commercialization possible, Fig. 19.2.

Fig. 19.2 Modular architecture



From a cost perspective, it quickly becomes apparent that commercial off-the-shelf (COTS) components provide the best cost model for the framework. With COTS, the prognostics systems designer minimizes electronic design as well as software design efforts. Instead, the designer is able to leverage development work already in play within the COTS suppliers' organization. Further, COTS systems typically provide a faster validation cycle as much of the hardware and software components have validation certificates.

This chapter examines core components of the prognostics system along with component interchange and COTS component options.

19.2 Modular Architectures

There are several publications promoting a modular architecture for condition monitoring and prognostics systems. One such standard is the ISO 13374 standard, Fig. 19.3, [1]. The ISO 13374 standard divides the condition monitoring and prognostics system into eight subsystems. These include data acquisition (DA) and data manipulation (DM) or data acquisition and signal processing. The ISO 13374 standard also calls out state detection (SD). State detection is often defined as the determination of deviation from normal or healthy operation. It can also be defined as determining the operational state of the machine, for example high speed operation and low speed operation.

Three prognostic functions in the ISO 13374 standard include Health Assessment (HA), Prognostic Assessment (PA), and Advisory Generation (AG). These three blocks perform the hard work of diagnostics, prediction, and information delivery. The outer two blocks on the left and right of the six middle blocks further define data base storage and retrieval, as well as visualization including technical displays and reporting. When following this model, the prognostics developer can save costs by foregoing the need to design these architectures. Further, when following a defined standard, it is possible to mix and match components from multiple commercial suppliers, each of which may specialize in a specific component area.

The University of Cincinnati, for example, takes a unique approach in adapting the ISO 13374 model by adding data filtering and sorting during the data acquisition (DA) step in the process, Fig. 19.4, [2]. The University recommends sorting data into operating regimes.

Operating regimes are distinguished by speeds, loads, and even mechanical failure modes. These regimes are identified during the data collection process. Data is labeled with the regime name, for example speed and load. Downstream categorization of information is made easy with regime tags made at the data collection point.

In either case, adaption of the ISO 13374 model to specific machine implementation provides modularity, flexibility, and promises to lower costs.

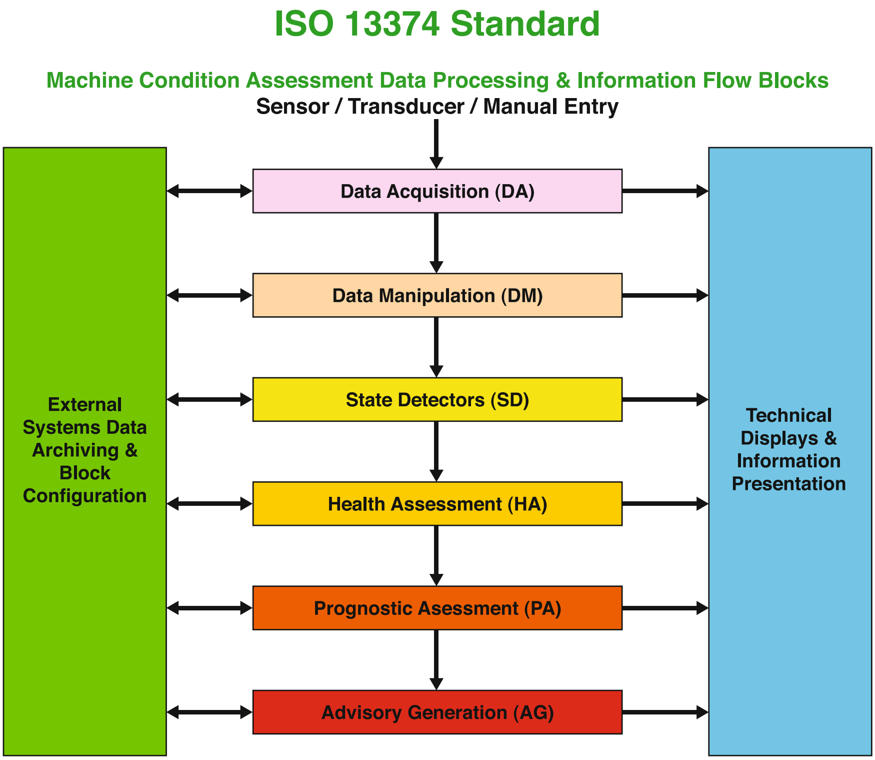


Fig. 19.3 ISO 13374 standard

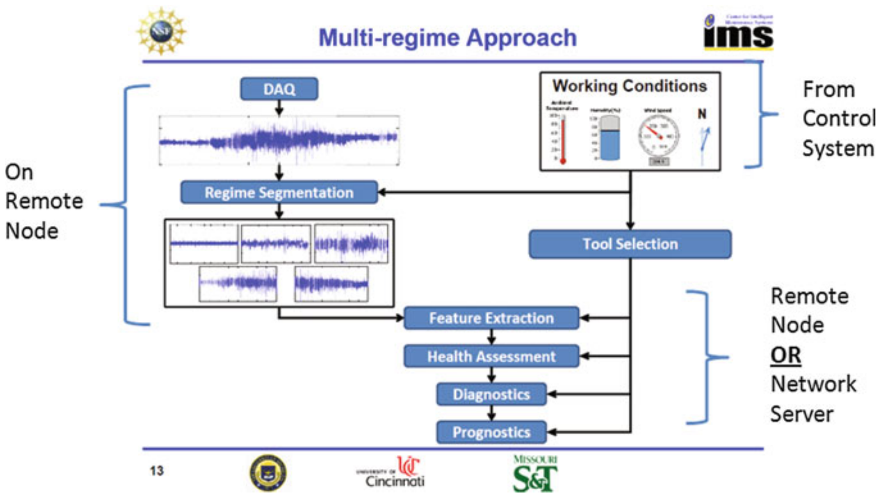


Fig. 19.4 IMS center multi-regime approach

As an example, consider a surface mine operation which employs swing shovels, drag lines, haul trucks, conveyors, crushers, and cogeneration gas turbines. Each of these critical assets has unique sensor data acquisition requirements, analytics specific to the varied mechanical components, and display or reporting needs. With a modular architecture, a common base platform of embedded processor and chassis is adaptable to each machine type using various sensory input cards within the chassis. In the embedded software, data recording triggering blocks are exchanged for each machine type to capture the right amount of sensory data at the right time. Embedded analysis is adjusted for embedded data reduction. However, the same communications modules, configuration screens, and so on are used across all assets at the mine location. At the maintenance facility, a common data base is used for all assets; however, prognostics modules are specific to machine types. Yet, with a common modular architecture, machine-to-machine comparisons are possible.

19.3 Data Acquisition Components of Prognostics

The data acquisition component of the prognostic system has the important role of correctly recording sensory information from the machine for downstream analysis and decision making. Data acquisition typically consists of sensors, signal conditioning, and an analog to digital converter; Fig. 19.5.

To obtain the best sensory information, many dynamic signals including vibration, need a wide dynamic range data acquisition device. Dynamic range is a measure of the data acquisition's systems ability to detect both strong and weak signals at the same time. In the case of vibration, it is important as many vibration sensors measure vibration from multiple machine components at the same time. In other words, high-amplitude low frequency vibration from unbalance is measured along with low-amplitude high frequency vibration from roller bearing and gears.

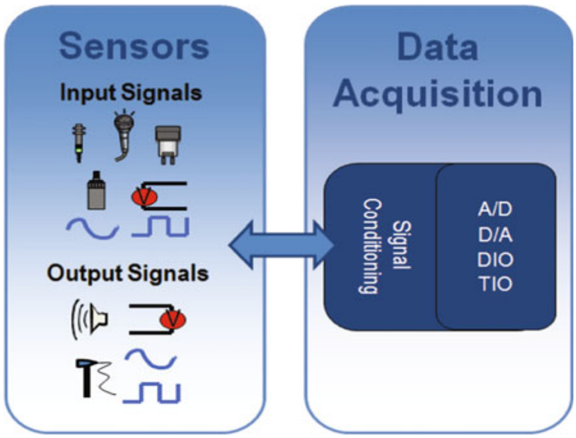


Fig. 19.5 Data acquisition system components

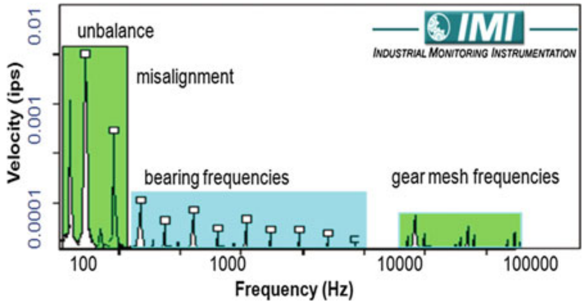


Fig. 19.6 Vibration frequency spectrum

A wide dynamic range data acquisition system, such as a 24-bit delta sigma analog to digital converter is beneficial in prognostic applications. The difference in amplitudes at various frequencies can be seen in the Fig. 19.6.

A high dynamic range then allows the single sensor to correctly digitize unbalance vibration, mechanical looseness vibration, bearing fault vibration, and even gear mesh vibration. In addition to dynamic range, there are several other factors for consideration in the data acquisition component. These include anti-aliasing filters, amplification, and sensor power. Finally, most 24-bit vibration acquisition hardware devices provide configurable IEPE 24 V power to power accelerometers, laser tachometers, dynamic pressure, and acoustic emission and microphone sensors.

To provide a data acquisition component of the prognostic system, there exist three core choices. First, it is possible to design the data acquisition system from the ground up, selecting analog to digital and signal conditioning semiconductor components, board manufacturers, embedded processors, programming languages, and so on. While this approach can lead to the lowest manufacturing cost for higher volumes, the electronic design domain expertise and time to market costs become prohibitive.

A second choice is to purchase a series of board level products following any number of standards such as PC-104. This choice offers the prognostics developer a range of suppliers to choose from and a range of generic processor and analog to digital boards that can work together. In most cases, however, the processor and analog boards have limited software facilities for analysis, data storage, and downstream diagnostics or prognostics. In other words, they provide a fundamental capability, primarily designed for control applications with limited dynamic range, and have limited software support. These products typically are designed to compete on price, with limited advanced features often needed for embedded or downstream analysis. The prognostics developer then must create and validate signal processing, filtering, and other related algorithms in addition to data storage and communications. This effort can become a significant software development effort.

A third choice is to build the prognostics system on a modular system designed for high fidelity sensor measurements with wide dynamic range, with a wide range

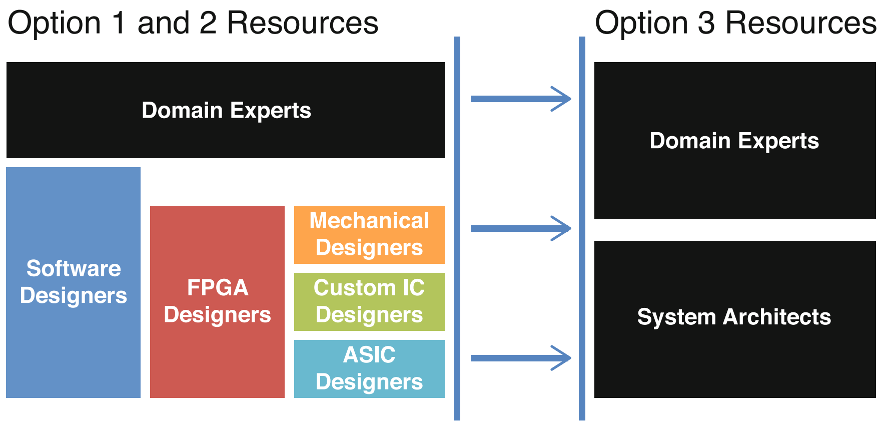


Fig. 19.7 Modular system development effort

of hardware certifications, and with a full featured signal processing library for downstream or embedded prognostic analysis.

The second and third options are differentiated by software development tools including mathematics as well as system certification activities. Figure 19.7 shows a comparison of complete custom (option 1) and modular system (option 3). There is considerable reduction in development effort required when using a instrumentation class modular system as the basis of a prognostics system.

Given, there is a data acquisition modular hardware platform in place, it is possible to adjust the sensory input capabilities of the system using modular hardware I/O to best match the machine for which the prognostic system will be used. A modular system allows analog input modules for a variety of sensor types to be selected based on the needs of the system. For example, vibration, temperature, speed, and process variable modules can be added (or removed) based on the measurements that best serve the prognostic system.

Finally, hardware should be industrial grade, meeting high standards for rugged environments. Key measures of rugged reliability include temperature, shock, and vibration. The prognostic systems designer should also consider supplier quality systems and hardware warranty. With a solid hardware framework, the prognostics developer is able to focus on the systems architecture, prognostics algorithms, and information delivery aspects of the system.

19.3.1 Data Recording Filters

Once the data acquisition system is chosen, it is prudent to determine what signal processing and data storage strategies should be embedded into the data acquisition system component of the prognostic system. On one end of the scale, all time series data from all sensors is continuously recorded. While this strategy insures no

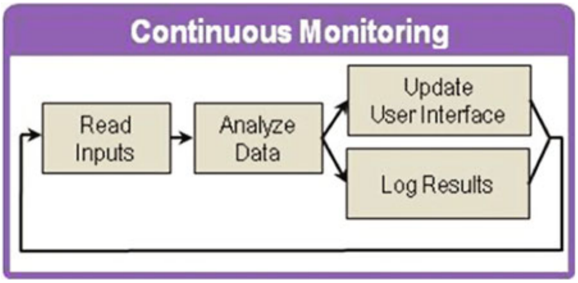


Fig. 19.8 Analysis driven data logging

loss of data from the machine, it burdens communications and downstream signal processing.

An alternative is to filter data to limit on board recording to just data that contains new information. This filtering us typically done by onboard analysis as illustrated in Fig. 19.8.

By analyzing data, it is possible to determine whether the data has changed. Examples of onboard analysis include statistical analysis, spectral analysis, speed, temperature, and process variables, etc. The prognostics system should be configurable to allow for deviation limits of sensory information to be used as data storage triggers. With this implementation in place, sensory data is recorded only when it has changed, on a periodic basis, or when an operator request has occurred. Further, the recording includes the condition which caused the data to be recorded. By recording a range of sensory metrics or features along with the recorded data, it is possible to sort the data downstream in the prognostic process. These sensory metrics, then allow the downstream prognostics functions to categorize operational and failure patterns of the same machine and similar machines.

When reviewing the capabilities of commercial off-the-shelf technologies for a prognostic system, it is prudent to consider software templates, routines, and facilities, etc. that allow for data filtering and data sorting. With the ability of the data acquisition system to filter and sort data, downstream prognostic consumers of the data are more focused and productive.

19.3.2 Data Storage Format Consideration

When considering data storage formats, it is best to leverage a technology or format that works well for the embedded data acquisition system, and provides rich descriptive capabilities for downstream prognostics analysis. A solid data schema defines data types including time waveform data, spectral data, alarm status, process data, and a range of sensory source information including machine asset and sensor asset information.

When the data sets are organized by sensor, mechanical component, etc., a view of related data sets is easily obtained. For example, opening a sectional view under

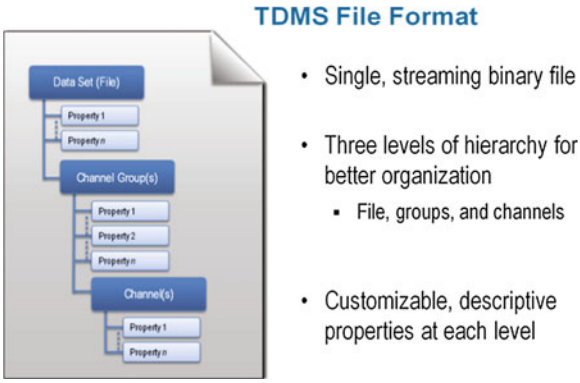


Fig. 19.9 Example embedded data recording structure

a roller bearing, one would see time series vibration data, temperature trends, vibration spectra, and oil particulate count trends, etc. All of the information is organized as sensory information related to the bearing. There are several ways to implement an embedded data storage capability which supports this rich data structure. These include common relational database structures and data structures specifically designed for embedded monitoring applications.

Figure 19.9 illustrates a data structure that is efficient in recording with its high speed streaming capabilities. It is rich in data descriptors with the use of data property strings for each data element or channel stored in the data file. In other words, information about the sensor, location, scaling factors, filtering, and mechanical component beneath the sensor, can be stored as labels that describe the time waveform recording. In addition, properties in the data file describe the conditions that caused the data file to be recorded, whether it be an analysis result threshold limit, a time limit, speed change, or operator request.

19.4 Signal Processing and Visualization

Signal processing functions operate on sensory data to extract features or measurements from data acquired from sensors placed strategically on the machine. Signal processing can occur in the data acquisition system, downstream on an engineering or database computer, or even across the Internet leveraging emerging cloud computing technologies. Signal processing plays a part in state detection, health assessment, and prognostic assessment steps in the complete prognostic system.

The choice of signal processing function is made on feature extraction needs, mechanical phenomenon indication desired, and domain expertise and preference of the prognostic system designer. It is important that the software development tools used to implement the prognostic system, offer a wide range of signal processing capabilities.

The IMS Center at the University of Cincinnati has added performance prediction, assessment, and diagnostic pattern matching as a supplement to advanced signal processing [3]. These capabilities operate downstream from embedded data acquisition by classifying extracted features into operating modes and failure modes. Health or performance assessment and prediction or prognostics assessment build on signal processing used in the data acquisition, data filtering and sorting, and feature extraction steps of the prognostic system steps. These additional steps, including logic regression, statistical pattern recognition, and self-organizing maps (SOM) provide tools for matching current system monitoring or data acquisition measurements with data driven models of system health and failure modes.

19.5 Conclusion

By leveraging commercial off the shelf technologies and a flexible modular architecture, it is possible to develop and bring to market a prognostic system that adapts to a wide range of machines, industries, and applications. The prognostics system developer is able to get to market rapidly and at less cost, than the alternative of developing components that are otherwise commercially available.

References

1. International Standards Organization (2003) Condition monitoring and diagnostics of machines—data processing, communication and presentation—part 1: general guidelines. ISO-13374-1:2003 Available from www.iso.org
2. Lee J (2009) Advanced prognostics for smart systems. Intelligent maintenance systems (IMS) introduction. Available from www.imscenter.net. p. 6–8
3. Intelligent Maintenance Systems (2007) Watchdog agent® documentation, center for Intelligent Maintenance Systems. Available from www.imscenter.net

Chapter 20
Investigation of Energy Consumption and Wear in Bypass Pneumatic Conveying of Alumina

B. Chen, A. A. Cenna, K. C. Williams, M. G. Jones and Y. Wang

Abstract Dense phase pneumatic conveying is critically dependent on the physical properties of the materials to be conveyed. However, many materials, such as alumina and coarse fly ash, which are highly abrasive, do not have dense phase conveying capacity. Bypass pneumatic conveying systems provide a dense phase capability to non-dense phase capable bulk materials. These systems also provide the capacity of lower the conveying velocity and therefore lower pipeline wear and lower power consumption occurs. The objectives of this work were to study the energy consumption and wear of bypass pneumatic transport systems. Pneumatic conveying of alumina experiments were carried out in a 79 mm diameter main pipe with a 27 mm inner diameter bypass pipe with orifice plate flute arrangement. High-speed camera visualizations were employed to present flow regimes in a horizontal pipe. The experimental result showed the conveying velocity of bypass system is much lower than that of conventional pipelines; thus, specific energy consumption in the conveying process is reduced. The service life of the bypass line has also been estimated.

B. Chen (✉) · K. C. Williams · M. G. Jones · Y. Wang
Centre for Bulk Solids and Particulate Technologies, Faculty of Engineering and Built Environment, The University of Newcastle, Callaghan 2308, Australia
e-mail: b.chen@studentmail.newcastle.edu.au

K. C. Williams
e-mail: ken.williams@newcastle.edu.au

M. G. Jones
e-mail: mark.jones@newcastle.edu.au

Y. Wang
e-mail: ywang11@uon.edu.au

A. A. Cenna
Mechanical Engineering, Faculty of Engineering and Built Environment, The University of Newcastle, Callaghan, NSW 2308, Australia
e-mail: ahmed.cenna@newcastle.edu.au

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_20,

20.1 Introduction

Pneumatic conveying systems provide flexibility and efficiency in their space requirement without generating environmental problems. Therefore, they are widely employed in the aluminum industry for its materials handling processes. There are two types of pneumatic conveying systems: dilute phase and dense phase. Dilute phase conveying process uses relatively higher mass flow of air (low solid loading ratios) which generates high conveying velocities. Material remains suspended in the conveying air throughout the conveying process. Unlike dilute phase conveying systems, Material is conveyed in much denser concentration of bulk solids at relatively low velocities through a conveying line. In dense phase mode, material will not be suspended in conveying air at least some part of the conveying line. Alumina is a very abrasive material. As a result, erosive wear of pipelines are significant in the process of dilute phase conveying of alumina. Dense phase conveying is critically dependent on the physical properties of the materials to be conveyed. Alumina does not have the natural dense phase capability to be transported in the conventional conveying pipelines. The solutions for these particulates to be conveyed in dense phase can be provided through the use of bypass pneumatic conveying system. Bypass systems can reduce conveying velocity thus lower power consumption and pipeline wear due to erosion. For this reason, bypass pneumatic conveying systems have been widely used in process industry for the last few decades.

Pavoni [1] summarized the technology of Fluidstat bypass systems and the advantages over conventional pipelines. Möller et al. [2] conducted alumina conveying tests to investigate the advantages of Turbuflow bypass system over conventional pipe line in the laboratory installation. The results showed that the pressure in the feeding vessel was more stable than the traditional system as segregation is prevented and the agglomerations are broken down. In order to investigate conveying characteristics of powder materials for a plant pneumatic conveyor, Ramkrishnan et al. [3] used a closed loop dense phase pneumatic conveying system incorporating a Turbuflow bypass system. Four different types of powder materials (cement, fly ash, pulverized coal, and raw meal) were transferred and the conveying energy consumption for a given volume of solids was assessed. The results showed the specific power consumption reduced with increasing phase density. Barton et al. [4] focused on overall performance of conveying alumina in bypass pneumatic conveying systems. Mathematical equations have been developed that describe how to determine the bypass pipe diameter and flute spacing.

There are two basic types of bypass systems; internal bypass and external bypass as shown in Fig. 20.1. One of the major problems with internal bypass systems is the unpredictable life of the bypass line due to wear. Material such as alumina is inherently abrasive by nature. As the bypass line is constantly in contact with these particles, pipeline is gradually worn out and reduces the effectiveness of

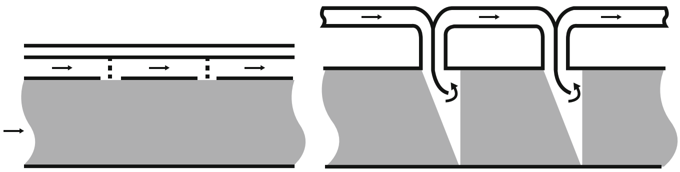


Fig. 20.1 Internal and external bypass pneumatic conveying systems

bypass conveying system. Once the bypass line is eroded, material will be conveyed in similar mode of the conventional pipeline. It is important to be able to predict the life cycle of the bypass line so that the material can be conveyed within the designated parameters and modes of flow. Understanding of the particles contacts plays a crucial role in determining the wear of the bypass line.

In order to obtain more detailed information of the solid phase during pneumatic conveying process, noninvasive flow visualization techniques have been employed in the past. An internal bypass pneumatic system was investigated experimentally and the internal flow pattern was visualized through electrical capacitance tomography (ECT) by Xu et al. [5]. The slug flow with a stationary bed revealed a wavy manner. No information on the bypass flutes area can be extracted from the ECT results. High-speed video camera (HSVC) has already been used to observe and analyze material flows in the pneumatic conveying pipeline through a glass section in some investigations, e.g., for the examination of slug and dune formation [6]. Chen et al. [7] employed high-speed camera to visualize flow regimes in a horizontal conveying of fly ash and explain material blockages inhibition in an internal bypass system. Cenna et al. [8] studied the flow structures in different bends while conveying alumina, sand, and fly ash [9].

Generally, research are carried out to investigate bypass pneumatic conveying systems with little attention being paid to compare the energy consumption and wear of pipelines of alumina in between bypass systems and conventional pipelines. Visualization of the operation of bypass pneumatic transport of alumina using HSVC is also studied in this paper.

20.2 Experimental Set-up and Procedures

The bypass pneumatic conveying test was conducted on a 6.5 m bypass system to obtain experimental data for the operation of a bypass system as shown in Fig. 20.2a. The material is fed into the system from the bottom of the feed bin which is a positive pressure blow tank. The discharge rate of materials from feed bin was controlled by adjusting the proportion of total air flow directly into the blow tank and changing the position of feed pipe. The receiving bin is mounted on load cell for measuring the mass flow rate. The main pipe is 80 NB and the bypass

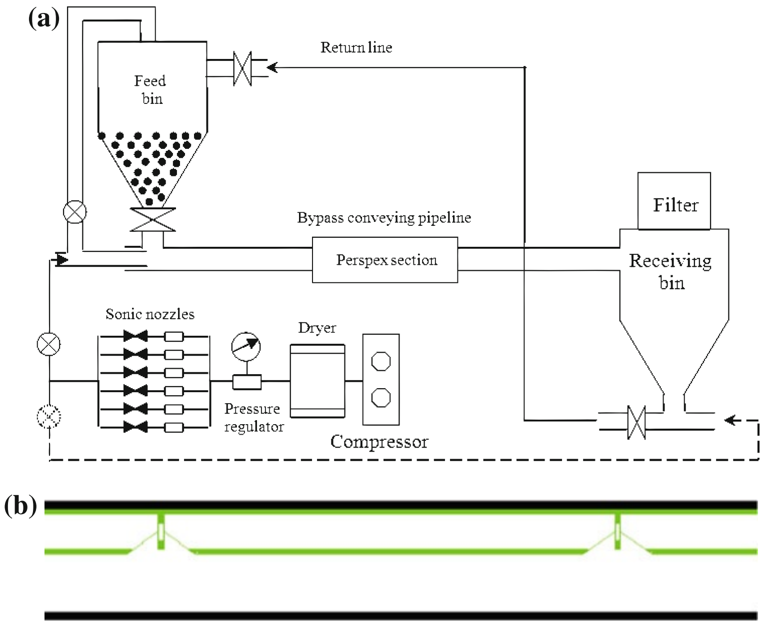


Fig. 20.2 Bypass pneumatic conveying test rig used in the current study. **a** Schematic diagram of the bypass system. **b** Arrangement of bypass pipe

Table 20.1 Bypass configurations

Main pipe ID (m)	Bypass pipe ID (m)	Bypass flute spacing (m)	Orifice plate diameter (m)	Angle of bypass opening (°)
0.079	0.027	0.4	0.007	45

pipe is 25 NB. The inner bypass pipe was designed using orifice plate approaches, which is shown in Fig. 20.2b.

The bypass configurations in terms of bypass flute sizes and flute spacing are shown in Table 20.1. In order to monitor real-time behavior of the system, pressure transmitters are used to measure the gauge pressure. A LabVIEW program is used to monitor and record the data at an acquisition rate of 50 Hz.

A glass section used in the main pipe allowed flow visualization by HSVC. HSVC Phantom 5 with 105 mm lens is used to obtain detailed information of solid phase flow behavior in pneumatic conveying pipelines. The image sample rate was up to 1,000 frames per second and the exposure time was 990 μs. Alumina has been used in the test and the detailed physical properties of alumina are shown in Table 20.2.

Table 20.2 Physical properties of material

Material	Mean diameter (μm)	Particle density (kg/m^3)	Bulk density (kg/m^3)	Permeability (x10^{-7} m^2/(Pa·s))	Minimum fluidization velocity (mm/s)
Alumina	76.7	4,088	1,003	3.6	17.7

20.3 Results and Discussion

20.3.1 Flow Visualization

Visualization of flow in dense phase pneumatic conveying of alumina in bypass system and conventional pipeline were conducted in this study. Figure 20.3 presented the flow patterns observed in conventional pipeline at different solids and air mass flow rates. The conveying parameters for these experiments are presented in Table 20.3, where SLR represents solids loading ratio. It has been demonstrated that for certain solids loading ratios alumina formed a stationery layer at the bottom and the particles flow on top of the layer at a higher velocity. With increased solid loading ratio, it is apparent that the stationery layer at the bottom becomes fluidized and changes the flow characteristics to moving bed flow. Alumina is a nondense phase material which cannot normally be conveyed in dense phase in the traditional pipeline. However, the study showed a remarkable characteristic of alumina conveying and a very high solids loading ratio. It is possible that this is a particular case of pneumatic conveying for a shorter pipeline without any bends. It would be worthwhile to investigate the effect of solid loading ratio on dense phase conveying of alumina in the straight pipe varying the length of pipeline in the future work.

Figure 20.4 presented the flow characteristics in the bypass pneumatic conveying of alumina over a range of superficial air velocities. The observation area including two bypass flutes is about 0.7 m in length, and the positions of the bypass flutes are shown in Fig. 20.4a. Figure 20.4b, c presented the flow characteristics of alumina conveying in a bypass system for different conveying parameters.

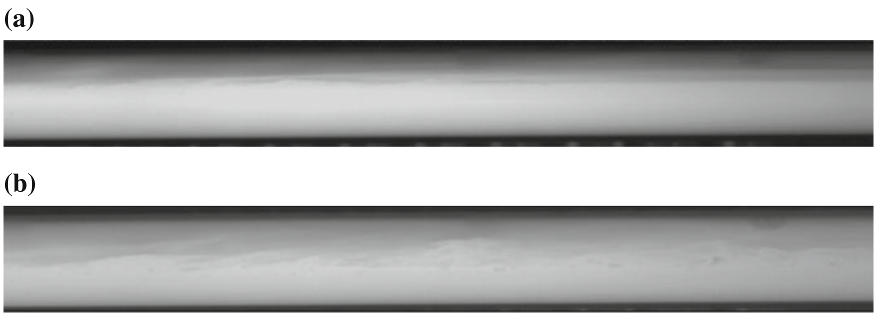


Fig. 20.3 High-speed camera visualization of pneumatic conveying of alumina in conventional pipe (Flow direction is from left to right). **a** Case 4. **b** Case 1

Table 20.3 Parameters used in conventional and bypass conveying of alumina

Conventional pipe Case No.	Ma, kg/s	Ms, kg/s	SLR	Vs, m/s	Bypass pipe Case No.	Ma, kg/s	Ms, kg/s	SLR	Vs, m/s
1	0.0681	2.87	42	15.2	6	0.0681	2.78	41	18.1
2	0.0595	2.10	35	13.5	7	0.0417	2.76	66	10.1
3	0.0556	2.29	41	12.2	8	0.0317	2.99	94	7.3
4	0.0681	0.85	12	23.4	9	0.0157	2.82	180	3
5	0.0595	0.44	7	21.4					

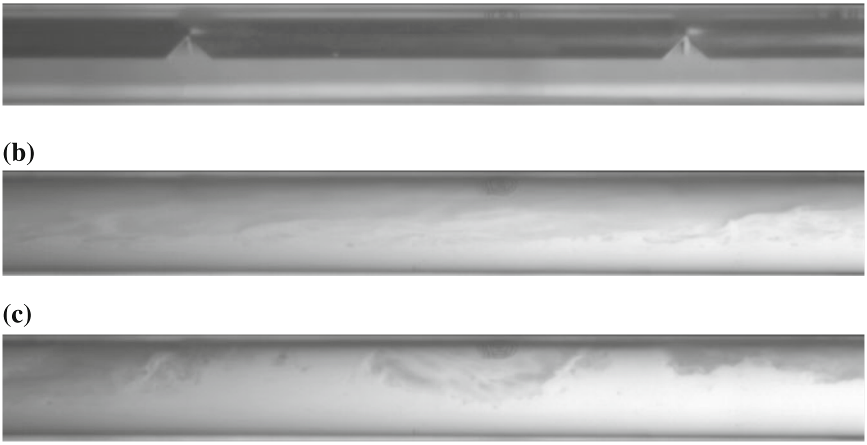
(a)
(b)
(c)


Fig. 20.4 High-speed camera visualization of pneumatic conveying of alumina in bypass pipe (Flow direction is from left to right). **a** Bypass flutes position. **b** Immature dune flow- case 7. **c** Slug flow-case 9

The conveying parameters including the solids loading ratios for by pass system are presented in Table 20.3. Figure 20.4b represents the flow pattern of alumina for a SLR of 66. Visual observation of the flow showed an immature dune flow. At higher solid loading ratios, the flow pattern reveals a slug flow. The slope at the front of the slug is steeper than the back of the wave as shown in Fig. 20.4c. The shape of the slug described by Konrad [10] is that the slope of slug front is less than that of the back. The difference of the phenomenon is caused by the properties of materials used in the tests. For the granular materials with high permeability in the Konrad's experiments, materials from stationary bed was picked up by the slug and then dropped behind the slug and finally form a stationary layer again. For fine powder in our tests, the rigid alumina slug was not formed as the gas cavity can be observed in the slug. In the bypass system, the pressure before the orifice plate in the bypass pipe is higher than that in main pipe as a result of orifice plate flow resistance. Therefore, air comes into main pipe and aerates the material continuously.

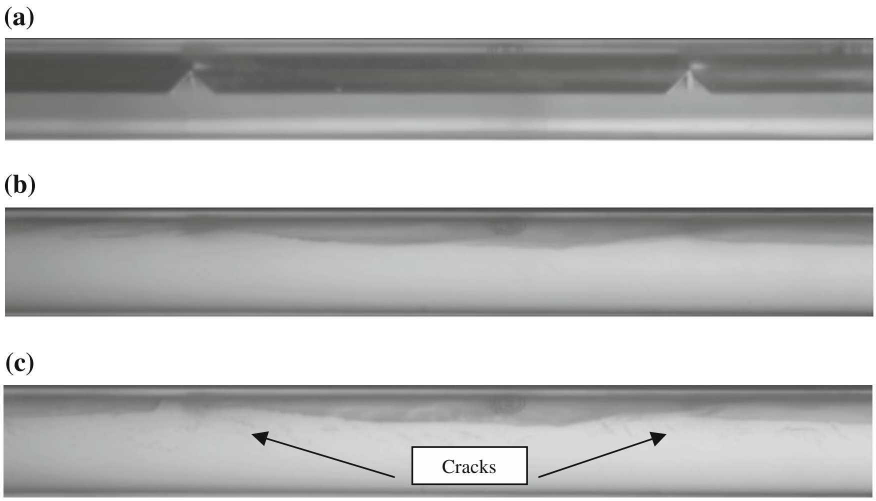


Fig. 20.5 High-speed camera visualization of aerating the material bed. **a** Bypass flutes position. **b** Stationary material bed. **c** Aerating the material bed

The visualization of the blockage clearance mechanisms is presented in Fig. 20.5. Figure 20.5b showed that the stationary alumina bed almost filled the whole cross-section of the main pipe. When the aerating process of alumina bed started in the bypass line, the cracks in the materials can be seen in Fig. 20.5c which was similarly observed in the experiments of conveying fly ash in horizontal bypass pipelines by Chen et al. [7]. The rigid full bore plug was not formed because air came out through the bypass openings into main pipe to penetrate into the material volume. The material is then fluidized slowly and blockage is cleared from the blocked location as seen in Fig. 20.5c.

20.3.2 Specific Energy

Specific energy is the energy needed to convey unit mass of material for a given pressure difference. It provides a simple measure for comparison between competing conveying methods. The specific energy consumption can be given by the Eq. (20.1) [11].

$$Specific\ Energy = 2RT\frac{M_a}{M_s}\ln\left(\frac{p_1}{P_2}\right) \tag{20.1}$$

where M_a is air mass flow rate (kg/s); M_s is solid mass flow rate (kg/s); R is universal gas constant (0.287 kJ/kg/K); T is absolute temperature (288 K); P_1 is conveying line inlet absolute pressure, P_2 is conveying line outlet absolute pressure.

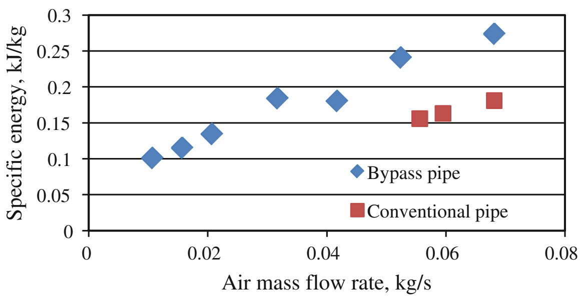


Fig. 20.6 Comparison of specific energy transportation between bypass system and conventional pipeline

The specific energy for both bypass system and conventional pipeline has been measured for the conveying tests carried out for visualization of flow structures. The results are presented in Fig. 20.6. For both bypass and conventional pipes, the results show that the specific energy decreases with the decrease of mass flow rate of conveying air. The difference in specific energy consumption between two conveying systems reduces with the decrease of conveying velocity. From the experimental results, it can be seen that the bypass pipeline consumes more energy than conventional system when using the same air mass flow rates. However, one of the advantages of bypass system is the ability to convey at lower velocities. With reduced velocity, the specific energy consumption is also reduced as supported by Möller et al. [2]. The pressure drop of bypass system is higher compared to conventional pipeline for same air mass flow rates, due to the increase of wall friction and flow resistance of a number of orifice plates. From Eq. (20.1), the specific energy rises with an increase of pipeline inlet pressure while other parameters remain the same. More energy is consumed in bypass system compared with conventional pipeline for same mass flow rates of solids and air.

20.3.3 Assessment of Wear of Bypass Line

One of the primary concerns of bypass system is the wear of the bypass line. For internal bypass system, there is no way to monitor the state of the bypass tube while in operation. On the other hand, if the bypass line failed in operation, blockage of pipeline is inevitable. The plant can operate without the bypass line at higher air mass flow rates whereby increasing the conveying velocity. This reduces the service life on the pipeline drastically. As a result, a predictive model for service life of bypass tube could save unscheduled breakdown keeping the reliable operation of the plant.

The particle velocity in the pipeline has been measure from the high-speed video of the flow. Figure 20.7 presented the measured average particles velocities around bypass pipe over a range of air mass flow rates in both conventional system and bypass pipeline. Based on the models developed for the assessment of service life of pneumatic conveying pipelines, the thickness loss of bypass tube has been

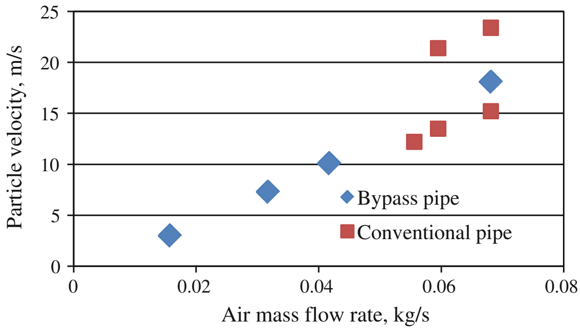


Fig. 20.7 Alumina particles average velocities

estimated. It has been estimated that for a wall thickness 3 mm of a bypass tube, it can make a hole in 2.5 years if the particle velocity 3 m/s whereas the bypass tube can be worn out in about 4 months if the particle velocity increases to 10 m/s. As the wear rate depends on the particle velocity together with other factors such as particle size, particle angularity, particle flux as well as the surface characteristics, the assessment of 'time to failure' for bypass tube needs a better understanding of the total operation of the system.

20.4 Conclusions

The bypass pneumatic experimental system was built with a main pipe of 79 mm in diameter and a bypass pipe of 27 mm inner diameter with orifice plate flute arrangement. The results showed that bypass system consume more energy than conventional system when using the same air mass flow rate due to the increased friction. The difference in specific energy consumption between two conveying systems reduces with the decrease of conveying velocity. The specific energy consumption is much reduced as the conveying velocity of bypass system is much lower than that of conventional pipelines. High-speed video camera visualization allowed the measurement of particle velocity in the pipeline. Based on the particle velocity, service life of the bypass line has also been estimated.

References

1. Pavoni G (2000) Fluidstat: low velocity conveying of coal-fired boiler fly ash. The third Israeli conference for conveying and handling of particulate solids. p. 130–136
2. Möller H, Pust J, Lubbe T (1985) Turbuflow: a pneumatic conveying system with economical power consumption. Bulk Solids Handl 5(4):789–794
3. Ramkrishnan T, Ramakoteswara R, Parameswaran M (1993) Experimental Studies on a Turbuflow System: A Pneumatic Conveying System with Economical Power Consumption. Adv Powder Technol 4(4):275–285

4. Barton S (1997) The Effect of Pipeline Flow Conditioning on Dense Phase Pneumatic Conveying Performance. [Ph.D. Thesis] Scotland, UK: Glasgow Caledonian University
5. Xu H, Liu S, Wang H, Jiang F (2002) Experimental Study on Wavy-flow Pneumatic Conveying in Horizontal Pipe. J Therm Sci 11(2):114–120
6. Konrad, K, Harrison, D, Nedderman RM, Davidson JF (1980) Prediction of the Pressure Drop for Horizontal Dense Phase Pneumatic Conveying of Particles. Proc. 5th Int. Conf. on Pneumatic Transport of Solids in Pipes. Bedford, England. P. 225-244
7. Chen B, Williams KC, Jones MG (2010) Experimental Investigation of Low Velocity Pneumatic Transport of Fly Ash in Bypass System. Advanced Materials Research 508:11–15
8. Cenna AA, Williams KC, Jones MG, Page NW (2006) Flow Visualisation in Dense Phase Pneumatic Conveying of Alumina. Proceedings of the World Congress on Engineering Asset Management. Gold Coast, Australia, pp 527–537
9. Cenna AA, Jones MG, Williams KC (2010) Wear of Pneumatic Conveying Pipelines: Flow Visualisation and Generation of Predictive Model. Bulk Solids Handling 30(4):190–195
10. Konrad K (1988) Boundary element prediction of the free surface shape between two particle plugs in a horizontal pneumatic transport pipeline. The Canadian Journal of Chemical Engineering 66(2):177–181
11. David D (2004) Pneumatic conveying design guide. 2nd edn. Elsevier Butterworth-Heinemann P. 247-252

Chapter 21
Condition Monitoring of Remote Industrial Installations Using Robotic Systems

M. G. Lipsett, J. D. Yuen, N. A. Olmedo and S. C. Dwyer

Abstract Distributed industrial systems present increasing risk as assets age. For systems such as pipelines and utility corridors, the cost of inspection to mitigate this risk needs to be controlled without compromising reliability. Major elements of remote inspection cost and effectiveness are the number of personnel, where they are located, inspection quality control, and travel time. One approach to reduce such costs is to use robotic systems for information gathering and preliminary feature extraction to detect anomalies and identify faults and their location. A combination of aerial and terrestrial robots can be deployed to cover the territory of interest and collect information necessary to extract features of interest in a timely manner. A system conceptual design is reviewed, and specific elements for a robotic mission to monitor the integrity of a pipeline and characteristics of a mine tailings structure are presented and discussed, with options for condition indicator data collection and feature extraction. Strategies are discussed for ensuring that the robotic inspection system itself has high reliability. Preliminary development and testing results for two prototype robotic systems are presented.

M. G. Lipsett (✉) · J. D. Yuen · N. A. Olmedo · S. C. Dwyer
Department of Mechanical Engineering, University of Alberta, Edmonton, AB, Canada
e-mail: Michael.lipsett@ualberta.ca

J. D. Yuen
e-mail: jdyuen@ualberta.ca

N. A. Olmedo
e-mail: olmedo@ualberta.ca

S. C. Dwyer
e-mail: scdwyer@ualberta.ca

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_21,

21.1 Introduction and Motivation

Industrial installations for power generation and resource extraction may be in remote locations. These installations include transmission and transportation systems along defined rights of way. Plants and transportation systems need to remain highly reliable, to meet customer needs and safety and environmental regulations. Risk mitigation for such distributed systems is based primarily on engineering controls, but monitoring during operations is used to ensure that the system remains within accepted limits of performance without violating constraints [1]. As technological assets age, the requirement to monitor may increase, because new failure modes can develop that require regular inspection to find and repair without compromising the entire system.

21.2 Previous Work on Remote Condition Monitoring

Measurements for condition monitoring are done remotely when access is hazardous or difficult for people, or when measurements are repetitive and costly to perform manually. Remote tooling has been employed for process monitoring and for diagnostics [2]. For example, robot-mounted cameras and simple sensors have been used for nuclear reactor vault monitoring and steam generator tube inspections [3]. Most industrial inspection tooling systems rely on teleoperation with human operation and observation.

Some demonstration prototypes have operated autonomously, including collecting condition monitoring information about the process and the robot itself. Remote sensing and sampling with unmanned rovers has been applied in rough terrain from the Amazon rain forest [4] to Mars [5]. Each application's requirements have driven the design of physical characteristics of the rover, payloads, and sensors. The Environmental Hybrid Robot for environmental monitoring has advanced the design and kinematic remodeling of rovers to safely and reliably monitor the Amazon rain forest [4]. Proposed space exploration rovers such as the ExoMars concepts include a wide range of sensors, sampling technologies and autonomous instrument placement [6, 7]. Sample collection concepts have been designed for lunar missions and implemented on autonomous rovers such as the Scarab Lunar rover prototype with a regolith core drill system [8].

Remote sensing has also been employed for agricultural monitoring, silviculture, and other applications over distributed regions. In precision agriculture, remote sensing can measure crop health to better target seed, fertilizer and pesticide densities, and delivery times. Studies have demonstrated the use of Unmanned Aircraft Systems (UAS) to aid in this process using techniques such as Normalized Difference Vegetation Index (NDVI) [9, 10]. Sensors used in these applications typically include thermal imagers and multispectral imagers in the visible and NIR bands. In particular, Berni et al. have shown lower cost, more favorable revisit times, and equivalent or better data quality when comparing a UAS platform to manned aircraft or satellite remote sensing [10].

Condition monitoring of utility corridors such as pipelines is an ongoing process required to demonstrate regulatory compliance [11]. Currently, these inspections are done using full-scale manned aircraft and ground inspection crews to take visual observations. It has been shown that spectroscopy using red edge analysis can detect vegetation stresses that can be linked to weeping hydrocarbon leaks [12]. Recently, the Schiebel S-100 Camcopter, a helicopter UAS, was used in Austria to perform powerline monitoring flights over several days. The high-definition images collected of the lines, poles, and isolators allowed technicians to work on the affected areas only. This was the first time in Europe that an unmanned helicopter was used in this manner [13].

21.3 Selecting Condition Indicators for Remote Monitoring of Industrial Installations, and Risk Mitigation for Remote Monitoring Systems

Tools commonly used to mitigate risk to industrial equipment are risk assessments, HAZOP, and Fault Modes and Effects Analysis (FMEA), which define the possible system failure modes the corresponding severity and probability. Condition monitoring is a control strategy to reduce the probability of an incident going undetected, thereby allowing measures to be taken to reduce the consequences of the incident. Some controls will require specific instruments and measurements for an industrial installation. For measurements in remote or hazardous locations, it may be possible to install dedicated instrumentation. In other cases, it is more effective to move the instrument package. In such cases, mobile and aerial robots can be a suitable option for condition monitoring.

The introduction of a complex robotic system into a remote monitoring regimen will affect the overall reliability of the industrial system. The robotic system must provide a remote monitoring solution with information of greater quality than current methods of information gathering (timeliness, accuracy, and additional features). The inspection system must have a higher reliability than the overall system, i.e., probability of successful monitoring must exceed the probability required to meet reliability specifications. A structured methodology is used to identify and control risks present in the robotic system, including assessing alternatives to control cost. This process also helps to define any required actions to deal with unacceptably high risk priorities.

The FMEA method is particularly well suited to mobile field robotics systems, as there are usually several operating scenarios or mission stages. For example, a mobile robot may have scenarios including deployment, data collection, retrieval, and fault recovery. All scenarios can be addressed to identify failure modes specific to a particular scenario. In addition, field robotic systems are subject to harsh, variable operating environments with a broad range of potential and possibly unexpected operating conditions. Identifying further condition-specific failure modes is useful in evaluating the system risks, their severity, and how controls can mitigate risk.

For a mobile field robotics system such as a UAS, there are a large number of subsystems and different mission stages with both minor and major failure modes. Examining a small-scale UAS powered by an internal combustion engine, major subsystems include the airframe and propulsion system, the autopilot system, as well as the payload system. Looking more closely at the propulsion system, one possible failure would be running out of fuel. A partial FMEA for this failure mode is shown in Table 21.1. The effects of this failure would be loss of thrust due to engine shutdown. This failure should then be evaluated for severity and occurrence. It is likely that the severity of the failure is near the middle of the scale, as it is likely a safe emergency recovery could be made engine-off with low threat to persons and property. Occurrence also depends on operating procedures (another form of control). This failure would present a much lower severity during mission stages such as start-up and shutdown that occur while on the ground. Detection of such a failure could easily be accomplished by monitoring engine speed and vibrations, or sound. Prevention may be accomplished through procedural controls, such as visually ensuring high fuel levels prior to takeoff and monitoring engine-on time during operation. In the event that the risk priority number (that is, the product of severity, occurrence, and detection) does not indicate low enough risk, additional controls should be considered, such as (in this case) a fuel level sensor to monitor fuel levels in real time, and a method for safely landing the aircraft before fuel is exhausted, employing a fault-tolerant control method to detect the anomalous state and to take a corrective action. Active control may require redundant sensing, control, and actuation systems to ensure that the control action can be executed.

Table 21.1 Sample FMEA analysis excerpt for a small-scale unmanned aerial system

Function or component	Fuel system
Operating scenario	Normal flight
Failure mode	Fuel depleted
Effects	Un-commanded engine stop in-flight
Severity (S)	7
Occurrence (O)	3
Current controls for prevention	Procedural controls: check fuel level prior to mission start, refuel if necessary; monitor throttle usage over time to predict fuel consumption
Current controls for detection	Onboard RPM sensor for engine status, audible detection within range
Detection (D)	6
Risk priority number (RPN)	126
Recommended actions	Install real-time fuel level sensor to improve monitoring accuracy

21.4 Mission Concepts and General System Requirements

Field-deployed robotic systems require relatively low cost to customize their design and mission for the specific task at hand. This allows for combined missions with various robots performing different inspections tasks (possibly simultaneously) as part of the same mission. A UAS could perform an initial mapping of the overall area to identify regions of interest for more detailed study. Then a ground robot can be dispatched to collect samples of these areas, which can also be used to fine-tune the classification methods for the aerial data.

21.4.1 Aerial

Manned aerial reconnaissance has been an attractive option for gathering remote sensing data due to the high resolution and survey size possible. With the small size and low cost of electronics and sensors today, it is feasible to field deploy small-scale UAS capable of delivering high quality, low cost remote sensing data. An added benefit of a field-deployed UAS is the potential for cooperation between the UAS and robotic ground vehicles. This could allow for online mission planning and reconfiguration during operation. For example, a UAS could provide aerial mapping to plan a trafficable route for the ground robot that passes regions of interest. Using direct measurements from the ground robot, the map can be updated and aerial measurements verified. The UAS could also act as a communications hub to greatly extend the communication range for the ground robot.

While UAS in general can range from hand-launchable to full sized jumbo jets, small-scale systems with dimensions not exceeding 3 m are of particular interest, because of their field portability, flexible launching (no need for a prepared runway), and reduced consequences in a crash. There are tradeoffs when designing a small scale UAS, such as payload capacity and endurance. These tradeoffs become more pronounced on smaller systems, and so it becomes important to tailor the airframe to the mission and the required payload.

For missions involving the harsh environments associated with industrial monitoring, there are a further set of specifications necessary to consider when creating a UAS for this mission. Many industrial installations operate in explosive environments or contain hazardous materials. Although the UAS may operate outside of the high-risk area, precautions must be taken to ensure any failure of the UAS does not cause failures of the installation. This is an important aspect to be considered in the FMEA for these systems.

A prototype UAS is currently under development to pursue research in these areas, as described below.

21.4.2 Terrestrial

Autonomous and teleoperated Unmanned Ground Vehicles (UGVs) have been integrated into a wide range of applications. The extensive need for robotic systems in civilian applications has pushed this technology into applications such as remote sensing, contact sensing, and surface and subsurface sample collection as part of industrial operations and investigations.

Mobile robot systems require certain specifications for reliable and safe autonomous operation. Autonomy of the robot depends on robust control loops and mission planning approaches. Path planning, trajectory generation, and mission specific control routines are used to navigate and operate payload. The robot's instrumentation and sensors allow for safe interaction with the environment. The reliability can be increased with redundancy of critical subsystems, machine monitoring, and fault detection. Safe operation can be achieved with proper understanding of the robot's operating conditions and environment. Communication with other equipment and workers operating in the same area is required to avoid interference and accidents due to poorly understood location and activity of the robotic equipment relative to other equipment and personnel in the area. Non-contact sensing (passive or active) can be used to detect possible hazards and modify trajectories. Manual override and emergency stops can prevent unplanned motions or collisions when the limits of the control loops are reached or the system is under unexpected operating conditions.

System requirements for the harsh environments of industrial monitoring must be addressed. Propulsion mechanisms and structural design of the rover must take into account the uncertainty of operating conditions such as uneven terrain and sinking. Proper control loops and sensors can help identify potential hazards and modify trajectories. Systems for industrial monitoring implement sensing and sampling technologies to minimize the risk and cost of manual sampling. Potential hazards for human operators can be minimized by introducing robotic systems to monitor dangerous industrial settings such as the oil sand tailings.

A prototype system has been developed at the University of Alberta for soft soil characterization and sample collection in rough terrain. The system concept includes a variety of features that will allow it to be integrated into industrial condition monitoring, such as the oil-sands tailing ponds. The key features proposed are the estimation and measurement of key soil parameters using an instrument package that includes a Fourier Transform Infrared Spectrometer (FTIR), an automatically controlled penetrometer, a shear-vane viscometer for subsoil measurements in soft soil, and sample collectors (surface sampler and subsurface core sampler). Simultaneous autonomous navigation and quantitative mapping of an area of soft ground will be possible using an adaptive traction controller, which estimates soil parameters during travel.

It is expected that a terrestrial system will cooperate with other systems for mission planning and reconfiguration. In an industrial setting such as the oil sand tailings, an aerial survey could deliver a preliminary map, including potential trouble areas, which the terrestrial system then uses for preliminary selection of sampling locations and trajectory generation of likely trafficable routes. Quantitative mapping of the terrain can be achieved as the terrestrial rover navigates over the surface and communicated its soil parameter estimation information in order to improve the aerial survey classification.

21.5 Development to Date on Applications of Interest: Industrial Pipeline Aerial Condition Monitoring and Soft Tailings Monitoring

Providing condition monitoring for industrial processes is important to ensure their safe and reliable operation. Although process controls can be used to identify many faults, there are hazardous conditions not readily detected without an external measurement device such as a UAS or UGV. Despite the added complexity of this system, by providing better condition monitoring it can lead to an overall more fault tolerant and reliable industrial process. Of particular concern is designing the additional monitoring system in such a way that failures do not lead to faults in the process itself.

Industrial hydrocarbon pipeline right of ways must be inspected periodically to ensure their safe operation. These inspections typically look for hazards such as weeping leaks, erosion, and vegetation or third party encroachment. In North America, this is usually performed as visual inspections by personnel aboard manned aircraft. There is opportunity to utilize UAS to perform this task using advanced sensors and change detection algorithms. Sensors could include thermal imagers to identify temperature differentials associated with leaks and spectral imagery to detect trace amounts of hydrocarbon or vegetation stress. Algorithms would allow for comparison of data collected at different times to highlight changes. This would allow for earlier detection of hazardous conditions, while reducing the workload and cost associated with pipeline patrolling.

In the summer of 2011, preparation began on developing a virtual and physical platform for UAS research. The physical platform will be used both as a sensor test bed and an autopilot control research subject. The virtual platform will enable software and hardware simulation prior to full integration in an effort to reduce failure occurrence rates. This platform will also be made available to other research groups at the university to foster cross-disciplinary cooperation and reduce start-up costs. The first prototype, based on a COTS Senior Telemaster model aircraft, is shown in flight in the photo in Fig. 21.1. It is equipped with a gas engine and autopilot system. Other aircraft designed by the team have been fitted with autopilot for autonomous operation, and have recorded geolocated, time-stamped video of ground objects using an automatically controlled pan-tilt unit.

The first payloads to consider will allow for electro-optical imaging in the visible spectrum with spectrometers to correlate shortwave infrared measurements

Fig. 21.1 Senior Telemaster for use as a UAS research platform

with ground parameters of interest. For example, the presence of hydrocarbons has been detected using adsorption features at 1,730 and 2,310 nm [14]. In mining applications, it has been demonstrated that airborne hyperspectral measurements can be used to predict the surface pH of pyrite mine tailings [15] and oil sands bitumen content can be estimated using broadband infrared reflectance [16]. This suggests that an extension of this work may allow the development of correlations for oil sands tailings trafficability with UAS spectral measurements. Future payloads may include thermal imagers, Light Detection and Ranging and Synthetic Aperture Radar. These instruments would greatly increase the utility of the UAS by providing greater spectral data and improving the rectification of measurements.

The condition monitoring of mining and tailings systems can be greatly enhanced by the use of robotic systems. An application of particular interest is the tailings deposits associated with bitumen production from surface mineable oil-sands in northern Alberta, Canada. Reclamation of tailings impoundments is only possible when the material has densified, because the varying consistency of deposits can cause earth-moving equipment to sink, endangering workers. To quantify the traversability of the terrain, key soil parameters need to be measured. The maximum shear strength that the wheel of a vehicle can impart on the terrain, max shear stress, can be calculated form cohesion and internal friction angle of the soil [17]. Iagnemma et al. [18] and Yoshida et al. [19] have proposed methods for estimating these parameters on board robotic systems. Robotic systems based on mission planning approached used for planetary exploration can be used to characterize oilsand tailings areas. The use of autonomous/teleoperated rovers to measure and estimate key soil parameters reduces the risk to workers as well as the time and cost of manual measurements. Low-risk measurements of key soil parameters can be used to identify hazards, zones for additional tailings work, and reporting for regulatory compliance [20]. Characterizing the wheel-terrain

Fig. 21.2 Testing of soil parameter estimation using Quadrover in dry sand

interaction will determine the mobility, and online processing will permit estimation of whether a path segment can be traversed.

The estimation of cohesion and internal friction angle can be achieved using a linearized model, presented on [5], from the slip-based traction model for wheel-terrain interaction presented by Bekker [17]. A least-squares or Kalman filter method can be used for parameter estimation from collected measurements including angular speed of the rover's wheels, linear velocity, normal load on the wheel, applied torque, and sinkage [5]. Preliminary tests with a prototype terrestrial rover system have been conducted in dry sand [20]. An indoor trial is shown in Fig. 21.2. The robotic platform used is based on a modified Quadrover, from Parallax, Inc [21]. This rover has been outfitted with various sensors and integrated circuitry for teleoperation and autonomous navigation. A ground station has been developed and implemented for remote operation and monitoring.

An FMEA is being conducted for each robotic system, and for a mission related to a tailings deposit mapping scenario.

21.6 Future Work

A design has been developed for a tooling payload for in situ soil testing with surface and sub-surface sample collection. This payload will be implemented on the rover, with a penetrometer. An FTIR point sensor will also be mounted on the rover to collect timestamped and geolocated hyperspectral data from the surface of the deposits. Field demonstrations will aim to collect measurements in real time during steady-speed driving on a soft tailings cell at an operating oilsands plant.

Results of the parameter estimation trials will be compared to those from conventional geotechnical tests to determine the industrial operational feasibility of the system. As well, alternative traction methods will be considered for very-low-strength deposits.

Small UAS are limited by endurance when compared to large unmanned aircraft. Normally addressed using design changes, future work with the UAS platform will include investigating the use of mission and path planning to reduce energy consumption and enhance endurance during flight. Within a similar planning framework, it is also possible to reduce risk during a mission. Energy consumption may be reduced through leveraging or avoiding local weather conditions (thermals, flying downwind), while risk reduction is possible by routing around known hazards (existing flight paths and airports, high population densities, environmentally or industrially sensitive areas).

There are currently several monitoring techniques applicable to remote industrial installations. An ongoing project aims to develop a quantitative design methodology to obtain an optimal remote monitoring strategy for a given situation. Examples of design parameters to consider are the targets and areas of interest, repeat timing required, and budget. These parameters will drive the selection and design of a monitoring system. A specific application of interest, such as oil sands tailings, will be used to validate the design methodology with the potential to use the UAS platform as a monitoring technique.

Work will be done to integrate a UAS with a rover into a geotechnical mission. The UAS can produce a preliminary map of surface conditions using features related to moisture and mineralogical characteristics that relate to shear strength. This map can then be used by the rover to collect more detailed hyperspectral imagery for soil characterization, and to collect samples, while assessing navigability during driving. A comprehensive FMEA will be part of system design and operations planning for each of these different demonstrations, to show how remote monitoring of industrial assets can improve reliability.

Acknowledgments The authors wish to thank the Natural Sciences and Engineering Research Council of Canada and the University of Alberta for funding support.

References

1. Bertling L, Allan R, Eriksson R (2005) A reliability-centered asset maintenance method for assessing the impact of maintenance in power distribution systems. IEEE Trans Power Syst 20(1):75–82
2. Rendon S, Fang J, Burghardt WR, Bubeck RA (2009) An apparatus for in situ x-ray scattering measurements during polymer injection molding. Rev Sci Instrum, vol 80(4). doi:10.1063
3. Chena Z, Yusab N, Miyab K (2008) Enhancements of eddy current testing techniques for quantitative nondestructive testing of key structural components of nuclear power plants. Nuclear Eng Des 238(7):1651–1656

4. Freitas G, Gleizer G, Lizarralde F, Hsu L, dos Reis NRS (2010) Kinematic reconfigurability control for an environmental mobile robot operating in the amazon rain forest. J Field Robotics 27(2):197–216
5. Iagnemma K, Dubowsky S (2004) Mobile Robots in Rough Terrain Estimation, Motion Planning, and Control with Application to Planetary Rovers, vol. 12 of Springer Tracts in Advanced Robotics. Springer Berlin/Heidelberg
6. Barnes D, Shaw A, Pugh S (2006) Autonomous sample acquisition for the exomars rover. In: Proceedings of the 9th ESA Workshop on Advanced Space Technologies for Robotics Automation
7. Woods M, Shaw A, Barnes D, Price D, Long D, Pullan D (2009) Autonomous science for an exomars rover-like mission. J Field Robotics 26(4):358–390
8. Wettergreen D, Moreland S, Skonieczny K, Jonak D, Kohanbash D, Teza J (2010) Design and field experimentation of a prototype Lunar prospector. Inter J Robotics Research 29(12):1550–1564
9. Suzuki T, Amano Y, Takiguchi J, Hashizume T, Suzuki S, Yamaba A (2009) Development of low-cost and flexible vegetation monitoring system using small unmanned aerial vehicle. ICCAS-SICE 2009 pp. 4808–4812
10. Berni J, Zarco-Tejada P, Suarez L, Fereres E (2009) Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. Geoscience and Remote Sensing, IEEE Transactions on 47(3):722 –738
11. CSA (2007) Oil and gas pipeline systems. CSA Standard Z662-07
12. van der Meijde M, van der Werff H, Jansma P, van der Meer F, Groothuis G (2009) A spectral-geophysical approach for detecting pipeline leakage. Int J Applied Earth Observ Geoinfo 11(1):77–82
13. AvioNews (2011) Uav helicopters. camcopter s-100 UAS: a strong impact in the civil market. http://www.avionews.com/index.php?corpo=see_news_home.hp&news_id=1131151&pagina_chiamante=index.php, June 2011.
14. Hörig B, Kühn F, Oschütz F, Lehmann F (2001) Hymap hyperspectral remote sensing to detect hydrocarbons. Inter J Remote Sensing 22(8):1413–1422
15. Zabcic N, Rivard B, Ong C, Muller A (2009) Using airborne hyperspectral data to characterize the surface ph of pyrite mine tailings. IEEE Workshop on Hyperspectral Image and Signal Processing: evolution in Remote Sensing pp. 1–4
16. Rivard B, Lyder D, Feng J, Gallie A, Cloutis E, Dougan P, Gonzalez S, Cox D, Lipsett M (2010) Bitumen content estimation of athabasca oil sand from broad band infrared reflectance spectra. Canadian J Chem Eng 88(5):830–838
17. Bekker MG (1969) Introduction to Terrain-Vehicle Systems. Michigan: University of Michigan Press
18. Iagnemma K, Shibly H, Dubowsky S (2002) On-line terrain parameter estimation for planetary rovers. Robotics and Automation. In: Proceedings. ICRA ’02. IEEE International Conference on, vol. 3, pp. 3142–3147
19. Yoshida K (2006) Terramechanics-based analysis and traction control of a lunar/planetary rover. Field Service Robotics 24:225–234
20. Olmedo N, Dwyer S, Yuen J, Lipsett M (2011) Soft soil parameter estimation and traction control of a gas powered rover for environmental monitoring of oil sands. In: Proceedings of the 23rd CANCAM, pp. 304–307
21. Parallax (2010) Parallax inc. http://www.parallax.com

Chapter 22
Systematic Design of Prognostics and Health Management Solutions for Energy Applications

M. Abuali, E. R. Lapira, W. Zhao, D. Siegel, M. Rezvani and J. Lee

Abstract Nowadays, energy has become a key issue in all sectors of industry. Analytical tools in the prognostics and health management (PHM) area are needed to transform energy and related data into actionable information. The IMS Center has developed corresponding solutions for energy generation, storage, and usage applications. The Watchdog Agent® toolbox techniques are applied in a systematic way in each application to address the development of advanced predictive tools for near-continuous uptime of energy generating assets; mobility readiness and safety for next-generation electric vehicles via the Smart Battery Agent; and the application of low-cost, nonintrusive predictive solutions using equipment energy consumption. In this paper, methodologies and case studies in the three categories are presented.

22.1 Introduction

As energy is a key performance factor of various industry systems, as well as a direct factor related to revenue and cost of the systems, interests in applying data analytics in energy related areas have been shown by original equipment

M. Abuali (✉) · E. R. Lapira · W. Zhao · D. Siegel · M. Rezvani · J. Lee
Center for Intelligent Maintenance Systems, University of Cincinnati, Cincinnati, USA
e-mail: abualima@mail.uc.edu

E. R. Lapira
e-mail: edzel.lapira@uc.edu

W. Zhao
e-mail: zhaowy@mail.uc.edu

D. Siegel
e-mail: siegeldn@mail.uc.edu

J. Lee
e-mail: jay.lee@uc.edu

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_22,

manufacture and users of various engineering asset. These areas include ones where energy is produced, stored or consumed, where different applications share the potential of utilizing existing prognostics tools and provide performance information related with system performance and cost or revenue.

The paper will present active research conducted by the IMS Center in implementing state-of-the-art prognostics and health management (PHM) techniques in a variety of systems related with energy, specifically three key scenarios/applications: (a) industrial facilities for monitoring and modeling of equipment energy consumption, (b) renewable energy applications for condition-based monitoring of wind turbines, and (c) mobility applications for the remaining useful life estimation of electric vehicle battery technologies. The paper will detail how predictive systems are designed for different energy applications based on similar core tools and techniques for signal processing, feature extraction and selection, health assessment and diagnosis, and performance prediction. In industrial energy applications, energy usage patterns are correlated with degradation behaviors of equipment and sub-components to further optimize overall energy management and utilization. In renewable energy applications, wind turbine operational and condition monitoring data is utilized to infer and predict future performance of wind turbine components (ex. bearings, gearboxes) and correlate the degradation patterns with the resultant power yield generated by the turbine. In mobility applications, predictive techniques are applied on batteries to realize "information-rich" energy-storage device with enhanced capabilities such as embedded service capabilities; remote smart monitoring for hard-to-access applications; and online management of batteries during their life-cycle to ensure mobility.

22.2 Energy Consumption Monitoring

A Precision Energy Management System (PEMS) has been developed to acquire energy consumption data in a non-intrusive manner [1]. Watchdog Agent® tools can then be applied to effectively monitor energy usage patterns over time and deviations from normal behavior caused by machine health or performance degradation or that can cause productivity loss can be detected before a pre-established threshold is reached. This PEMS system can be implemented at multiple levels (Fig. 22.1):

Power sensors can be installed at component level and machine level to measure energy signal at both levels. The energy signals are then segmented based on machine cycles, decoupled based on component motions and analyzed to correlate with component degradation, machine performance, and process status. The framework of energy-related prognostics analytics is presented as figure below (Fig. 22.2):

In one case study, continuous machining (milling) cycles with multiple cutting tools were conducted in a test-bed provided by industry. Dimension measurement was taken per each part after it is manufactured. Energy signal (in kilowatt) was

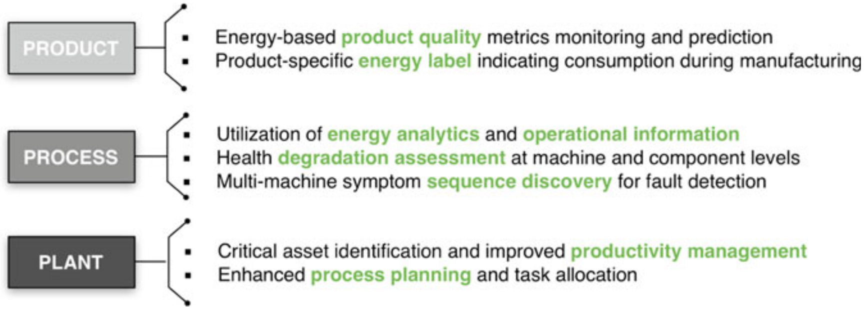


Fig. 22.1 Key values of PEMS with multiple objectives

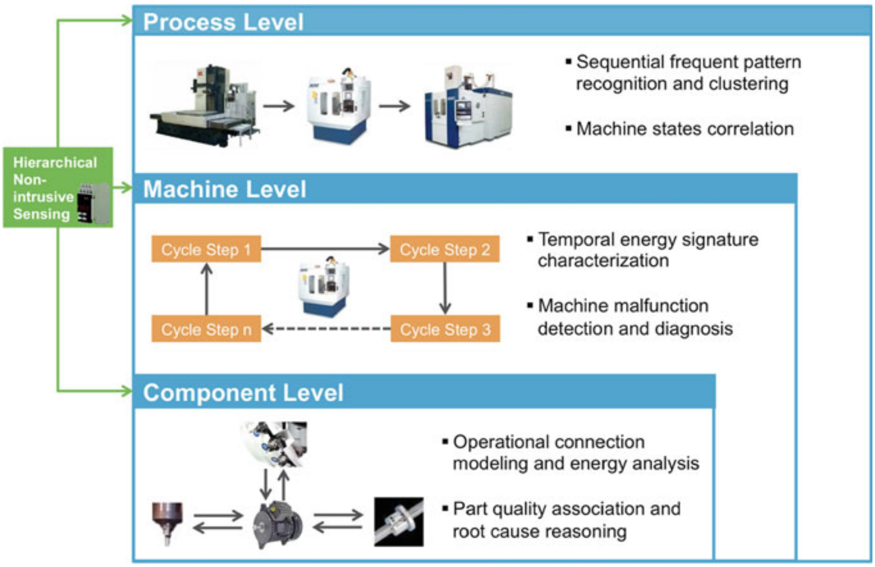


Fig. 22.2 Framework of energy-related prognostics analytics at different levels

taken from the milling machine and representative features are extracted from each milling cycle. The deviation of the feature distribution is modeled and correlated with the measured product quality parameter. Due to the high similarity between the two trends (Fig. 22.3), predictive method can be used to prevent part with part quality.

In another case study related with energy consumption analysis for industry system [2], a highly automated manufacturing line is studied where a power sensor is instrumented on each machine. Both effective power (kW) and reactive power (kVAr) were measured, where reactive power is considered to be able to reflect machine operation change and used to segment all data into machine cycles and cycle steps. Signal processing tools are then applied to characterize data signature

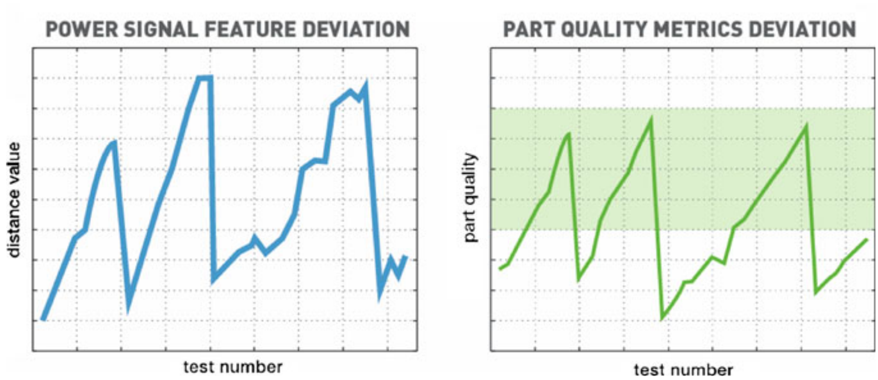


Fig. 22.3 High correlation between power signal deviation and part quality deviation

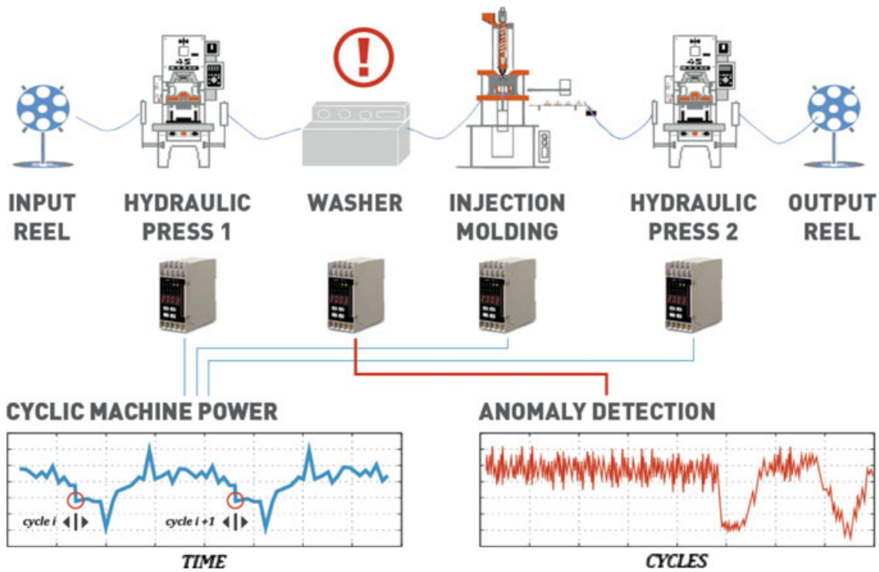


Fig. 22.4 Cyclic machine power signal is used to detect machine anomalies

by extracting features, and pattern recognition tools are used to detect feature distribution change and corresponding machine anomaly (Fig. 22.4).

22.3 Wind Turbine Prognostics

In this section, an emerging renewable energy area is investigated and the capabilities of prognostics tools in wind turbine monitoring are presented.

Mechanical failures and performance dropdown will decrease turbine production significantly and bring considerable issues in maintenance and logistics. Based on extensive literature survey, critical components that affect turbine downtime include drivetrain components like gearbox, shaft, and generator. Although much research effort has been focused on applying condition based monitoring techniques in this area, there are some remaining issues and unmet needs:

- There is a lack of performance metrics that compels maintenance activities;
- Wind turbine operate under dynamic operating conditions, which increase the complexity of system behavior;
- Reconfigurable and systematic PHM platforms are needed to improve overall wind turbine reliability.

Therefore, a smart wind turbine PHM platform is proposed by Lee et al. [3] (Fig. 22.5):

Supervisory Control and Data Acquisition (SCADA) system is usually a default monitoring system for wind farms when turbines are installed. Output variables and operation condition variables are used to model the continuous multi-regime behavior of wind production, based on the design of wind profile (power output vs. wind speed). A performance assessment and prediction method is then used to compare recent behavior with baseline behavior as well as to predict the deviation. This Global Health Estimator (GHE) is used to monitor the power generation capability of one turbine unit; therefore it is related with the prediction of when the turbine revenue will drop below a breakeven level. Upon the occurrence of the event, a Local Damage Estimator (LDE) is trigged to analyze vibration data from

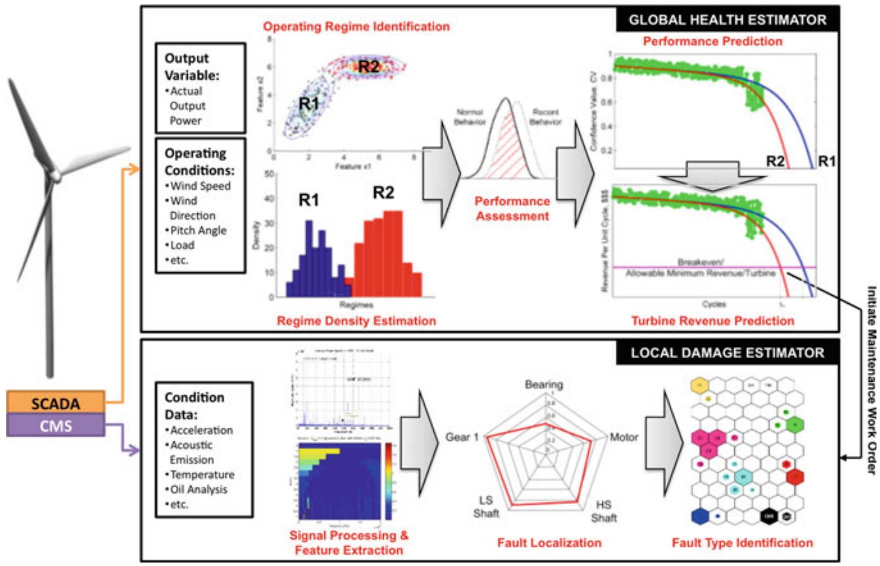


Fig. 22.5 Systematic framework for smart wind turbine PHM

Condition Monitoring System (CMS) and extract expert features. Drivetrain faults then can be detected, located, and diagnosed [4].

In the case study for this section, SCADA data of approximately 26 months is analyzed with multi-regime health monitoring tools including Gaussian Mixture Model (GMM) and Self-organizing Maps (SOM). The turbine experienced several major downtime durations due to unknown failures. Relevant variables are selected from over a hundred of them, and instances are filtered based on expert knowledge of turbine production and control mechanism. Data is then segmented into weeklong intervals and the distribution of the data is evaluated.

GMM (upper chart of Fig. 22.6) shows a more gradual degradation while SOM shows an abrupt increase of Minimum Quantization Error (MQE) distance. Therefore, GMM is better for degradation modeling and SOM-MQE is better as an anomaly detection method [5].

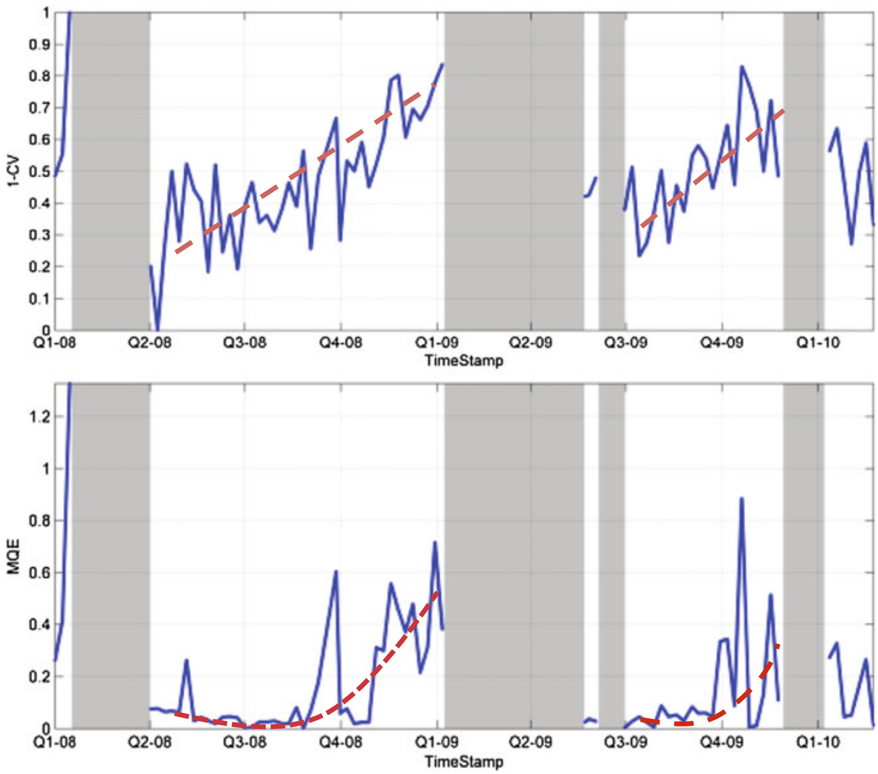


Fig. 22.6 Degradation of wind turbine energy generation performance

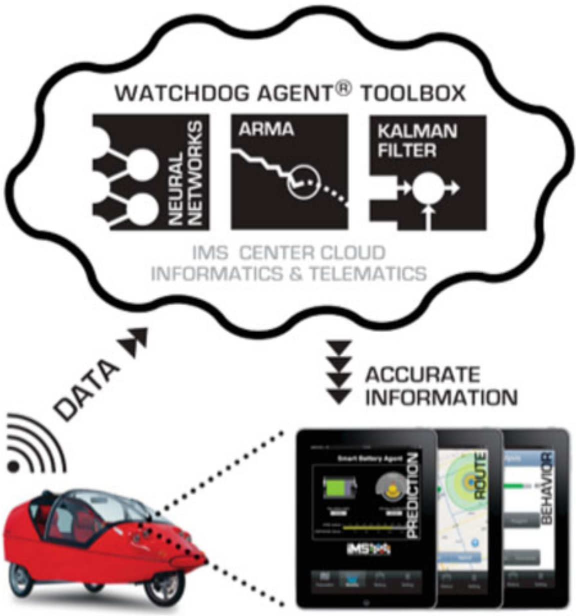


Fig. 22.7 Cloud-based informatics and telematics platform

22.4 Smart Battery and Mobility Systems

The Smart Battery techniques help to transform batteries into information-rich energy-storage devices with enhanced functionality [6], including: embedded service capabilities, remote smart monitoring, and online management of batteries to ensure mobility of electric vehicles (EV).

Mobility refers to the distance that an EV can travel, which is measured accurately via the cloud-based Smart Battery Agent by incorporating battery health

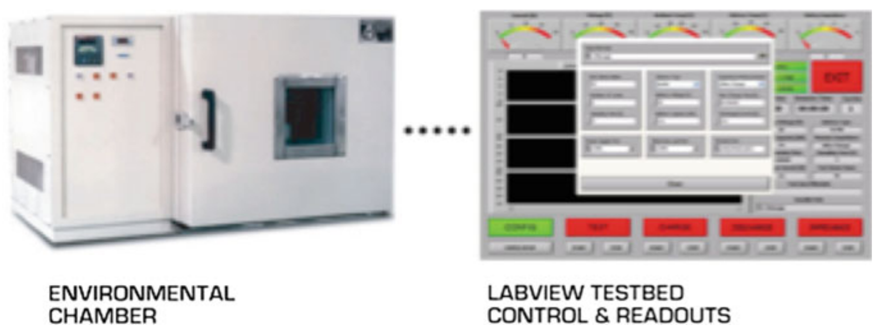


Fig. 22.8 Life-cycle testing for development of smart battery techniques

parameters, as well as environmental and user data, to alleviate range anxiety which has been a major impediment to the adoption of EV (Fig. 22.7).

The smart battery techniques are developed based on battery data from both in-house test-bed and in-field vehicles. For former scenario, life cycle testing is conducted where batteries are charged and discharged for large number of cycles until the capacity of the battery is decided to be below a pre-defined threshold. Cyclic data, including voltage, current, temperature, is then analyzed and appropriate degradation model is built (Fig. 22.8).

22.5 Conclusion

In the paper, development and application of PHM techniques in energy production, storage, and consumption areas are presented. For energy consumption monitoring, a PEMS is designed to utilize power signal and interpret machine performance at multiple hierarchy levels. In energy generating applications, critical asset such as wind turbines are monitored through analyzing their production performance. As for energy storage device, smart battery techniques are developed to extract health information of battery and provide advisory information of vehicle mobility in EV application. I acknowledge the work of our IMS researchers: Wenyu Zhao, Edzel Lapira, and Mohammad Rezvani.

References

1. Abuali M, Zhao W, Lee J (2011) Systematic framework for non-intrusive prognostics and health management (NI-PHM) using energy analytics. In: Proceedings of the 24th international congress on condition monitoring and diagnostics engineering management (COMADEM), Stavanger, Norway, 30 May–1 June 2011
2. Zhao W, Abuali M, Lee J (2011) A comparative study of power-based machine health assessment models. In: Proceedings of the IEEE international conference on prognostics and health management, p 1–6
3. Lee J, Lapira E, Zhao W (2011) An advanced prognostics and health management (PHM) framework for wind turbines: methodology and case study. In: Proceedings of the 2011 China wind power conference, Beijing, China
4. Al-Atat H, Siegel D, Lee J (2011) A systematic methodology for gearbox health assessment and fault classification. Int J Prognostics Health Manage Soc 2(1):16
5. Lapira E, Siegel D, Zhao W, Brisset D, Su J, Wang C, Abuali M, Lee J (2011) A systematic framework for wind turbine health assessment under dynamic operating conditions. In: 24th international congress on condition monitoring and diagnostics engineering management (COMADEM), Stavanger, Norway, 30 May–1 June 2011
6. Rezvani M, Abuali M, Lee S, Lee J et al. (2011) A comparative analysis of techniques for electric vehicle battery prognostics and health management (PHM). SAE technical paper 2011-01-2247. doi:10.4271/2011-01-2247

Chapter 23
Physics of Failure Based Reliability Assessment of Electronic Hardware

S. Menon, Elviz Georgea, M. Osterman and M. Pecht

Abstract Electronic products are continually developed as well as redesigned to improve performance and enhance reliability. These products are required to meet reliability requirements of the customer. Traditionally, reliability of products has been ensured by physical testing, which can be time consuming and result in design change iterations delaying product introduction or delivery. To reduce time to market, improve reliability, and product development costs, simulation techniques that can replicate physical tests can be used. Simulation must include consideration of product response to anticipated use and test conditions, as well as identification of failure mechanisms and failure sites including estimated time to failure. In this paper, virtual qualification using simulation assisted reliability assessment (SARA®) approach is introduced. Failure mechanisms of electronic products are discussed and methods for estimating time to failure are provided. Finally, a case study on the life assessment of a commercially available flash drive is presented. The predicted results are compared to experimental test results to determine the accuracy of the approach.

S. Menon (✉) · E. Georgea · M. Osterman · M. Pecht (✉)
Center for Advanced Life Cycle Engineering (CALCE), University of Maryland, College Park's, USA
e-mail: smenon85@calce.umd.edu

M. Pecht
e-mail: pecht@calce.umd.edu

M. Osterman
e-mail: osterman@calce.umd.edu

M. Pecht
Center for Prognostics and System Health Management, City University of Hong Kong, Kowloon, Hong Kong

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_23,

23.1 Introduction

From cell phones and mp3 players to personal computers to automated banker teller systems, electronic equipment has become ubiquitous with modern life. Depending on cost and application, users of equipment have an expectation of product life expectancy. In addition, equipment manufacturers have a certain life expectancy for their products and provide product warranties to assure prospective buyers that the product will meet a minimal life expectancy. The estimation of life-expectancy of a product under specified load conditions will assist the manufacturers in determining the warranty period, schedule of maintenance, replacement, and so on.

One of the most commonly used methods for electronic product qualification and reliability assessment is to conduct physical testing. Physical tests are based on the expected field conditions. These tests replicate stresses in the product that will occur under use. To shorten tests, stress levels are often elevated beyond expected use levels. When the physical failures that result from the test can be correlated to the stress level, the time to failure in test can be related to the time to failure in the field. This process is done by establishing the acceleration factor for the test. Tests that make use of elevated stress levels with established acceleration factors are accelerated tests. Guidelines to conduct accelerated tests and to determine acceleration factors are provided in various national and international standards such as IPC, JEDEC, MIL, TELCORDIA, JIS, and so on. However, tests may be conducted at elevated stress levels without established acceleration factors. In these cases, the object may be simply to have the product survive a specified test period with design changes being made until product can survive the prescribed test. Since testing is carried out under elevated load conditions, failures which may not be representative of actual failures in field conditions may arise. Therefore, elevated stress tests need not address the fundamental physics involved in the failure of electronic products. Further, survival of the test does not necessary relate to field survival. An accurate evaluation of product reliability requires the application of scientific principles and an understanding of the physical processes that produce failures [1, 2].

Computer simulation has been used to predict the behavior of electronic products and determine the strain/stress arising in electronic products during various loading conditions. The determination of strain/stress in electronic assemblies assists in understanding the physics of failure in electronics products. The use of computer simulation reduces the life cycle costs and product development time by eliminating actual physical tests. General-purpose simulation software, such as Ansys, Abaqus, Flotherm, and IcePack, have been used to evaluate stresses and strains in electronic circuit assemblies [3, 4]. Even though simulation takes lesser time than accelerated tests, accurate determination of strain/stress requires high mesh density and careful analysis of regions of highly concentrated stress, which in turn requires high processing capabilities. Furthermore, computer simulations do not reveal product reliability. For instance, none of the general-purpose simulation software provides the life expectancy of the electronic products under different loading conditions.

In order to estimate the life expectancy of electronic products under various loading conditions, physics of failure (PoF) models can be utilized [2, 5, 6]. PoF models can be broadly classified as strain-based, energy-based, and damage-based. PoF models provide analytical relationships between the cycles to failure and strain/energy/damage. The strain and/or energy developed in electronic products during loading can be estimated using general-purpose simulation software. To evaluate the accuracy of the predicted results, the estimated cycles to failure from PoF models should be compared and calibrated with experimental test data. The objective of the paper is to provide an overview of Simulation Assisted Reliability Assessment (SARA®), a methodology that incorporates PoF based principles and software to perform qualification and assess the life expectancy of an electronic product.

23.2 Simulation Assisted Reliability Assessment

Simulation Assisted Reliability Assessment (SARA®) is a methodology developed by CALCE to provide a simulation environment for rapid failure assessment of interconnects in printed wiring assemblies [7–10]. However, failure assessment using SARA is not limited to fatigue failures occurring in interconnects. The SARA methodology can also be used to assess failures occurring as a result of overstress loads that may lead to fracture. Models for electromigration, corrosion, dielectric breakdown, die attach fatigue among others are also included in the simulation environment. The methodology can also be used for assessing the electrolytic life of capacitors and risk of tin whisker growth in electronic assemblies.

The application of this methodology thus provides the manufacturers and designers an opportunity to assess time to failure, make design improvements, and design accelerated tests for various applications. The method emphasizes an understanding of the physical processes that produce failure in electronic systems. The method involves quantitative modeling of the expected failure mechanisms and obtaining quantitative damage metrics for estimating the remaining life of the components on the printed wiring board. The physics-of-failure stress and damage accumulation analysis involves using the material properties and geometry of the product and the measured life cycle loads to assess the dominant failure mechanisms. Based on a load-stress simulation, the physics-of-failure (PoF) damage models give an estimation of the accumulated damage for the product in its life cycle environment. The assessment of life expectancy under anticipated life cycle loading conditions is referred to as the virtual qualification (VQ) process.

23.3 Virtual Qualification

Virtual qualification (VQ) is a simulation-based methodology used to determine if a product meets its specified life requirement. The process is based on identification of potential failure mechanisms and sites followed by the evaluation of time to failure for based on these identified failure mechanisms and sites [11]. VQ enables the determination of the acceleration factor for a given set of accelerated test parameters, and to determine the time-to-failure corresponding to the identified failure mechanisms. VQ requires significantly less time and money than accelerated testing to qualify a part for its life cycle environment.

Each failure model in the VQ comprises of a stress analysis model and a damage assessment model. The output is a ranking of different failure mechanisms, based on the time-to-failure. The stress model captures the product architecture, while the damage model depends on a material's response to the applied stress. This process is therefore applicable to existing as well as new products. The objective of virtual qualification is to optimize the product design in such a way that the minimum time-to-failure of any part of the product is greater than its desired life. Although the data obtained from virtual qualification cannot fully replace those obtained from physical tests, it can increase the efficiency of physical tests by indicating the potential failure modes and mechanisms that the operator can expect to encounter. Ideally, a virtual qualification process will involve identification of quality suppliers, computer-aided physics-of-failure qualification, and a risk assessment and mitigation program. The process allows qualification to be readily incorporated into the design phase of product development, since it allows design, test, and redesign to be conducted promptly and cost-effectively. It also allows consumers to qualify off the-shelf components for use in specific environments without extensive physical tests. The software has been implemented to provide the following capabilities (Fig. 23.1).

- Thermal analysis
- Vibration analysis
- Failure assessment.

  - Identification of weak-links in the assembly design
  - Ranking of potential failure mechanisms
  - Estimation of operational lifetime, based upon identified failure mechanisms.

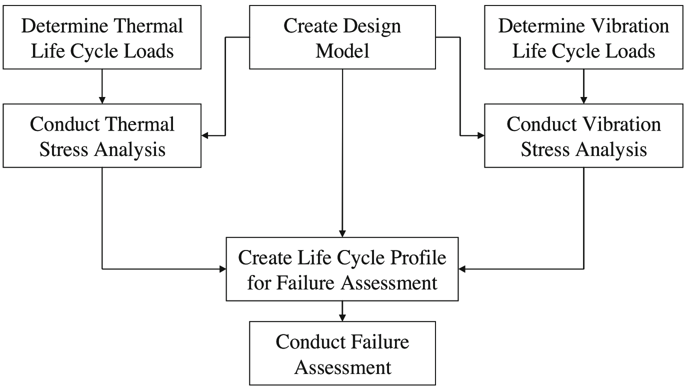


Fig. 23.1 Software capabilities

The software is meant to provide a rapid assessment. As such, the software has limited capability for conducting parametric studies and is not meant to replace sophisticated general-purpose analysis tools. It is rather designed to evaluate nominal designs. The steps involved in the VQ (Fig. 23.1) process include:

- Creation of design model
- Life cycle load characterization
- Load Transformation
- Failure risk assessment
- Ranking Potential.

23.4 Case Study

In order to demonstrate the application of the virtual qualification approach, the process was applied to a commercially available flash drive (shown in Fig. 23.2). The circuit card assembly for the flash drive is a double layer board that includes packages such as a TSOP, PQFP, capacitors, and resistors. To evaluate the performance of the virtual qualification approach, the circuit card was subject to a thermal cycling test. The times to failure for the components on the assembly were calculated using PoF models. Thermal fatigue is modeled using the Coffin-Manson based equation. In this model the fatigue life is related to the strain range in the solder joint during one cycle. The median cyclic fatigue life is given by:

$$N_f = \frac{1}{2}\left(\frac{\Delta\gamma}{2\varepsilon_f}\right)^{1/c}, \tag{23.1}$$

where $\Delta\gamma$ is the shear strain range at stress relaxation, ε_f is called the fatigue ductility coefficient in shear. Engelmaier [12, 13] published an expression for the fatigue strength exponent c based on cyclic mean temperature and the frequency of cyclic loading (dwell times at high temperature):

$$c = c_0 + c_1 T_{sj} + c_2 \ln\left(1 + \frac{360}{t_{dwell}}\right) \tag{23.2}$$

where T_{sj} is the mean temperature during a cycle, and t_{dwell} is the duration of the high temperature dwell (Fig 23.2).

The dimensions of the board and components were calculated for input in the PoF models. Material properties for the packaging, leads and solder joints were also found for each of the components on the flash drive to be used as input for the modeling. The temperature profile for the test was a minimum temperature of −55 °C to a maximum of 125 °C with a ramp rate of 10 °C/min and 15 min of dwell time at both extremes was included into the model. The analysis was then carried out to estimate the cycles to failure.

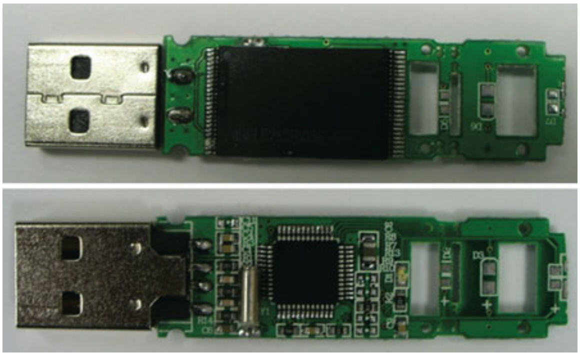


Fig. 23.2 Printed wiring assembly in a flash drive

Ten samples of the flash drives were also placed in an accelerated temperature cycling environment. The flash drives were subject to the same temperature profile as used in the simulation. In order to determine if the flash drives were still functioning, a functional check of all the samples was done once every 20 cycles using a memory check software. The software performs a performance check and a scan of the memory for defects. The test was terminated at 1,200 cycles with 3 samples surviving and the cycles to failure were plotted as a 2 parameter Weibull distribution (shown in Fig. 23.3).

Using the PoF based simulation, the mean time to failure for the TSOP was found to be 472 cycles and the PQFP was found to be 115,000 cycles. In order to compare the PoF analysis with the accelerated test results, the mean cycles to failure based on the test results using the 2 parameter Weibull distribution with 90 % confidence bounds was extracted. The mean cycles to failure was found to be 644 cycles with 365 cycles as the lower bound and 1,137 cycles as the upper bound. This showed that the PoF analysis results for the TSOP fell within the bounds of the test results. This was further corroborated using failure analysis.

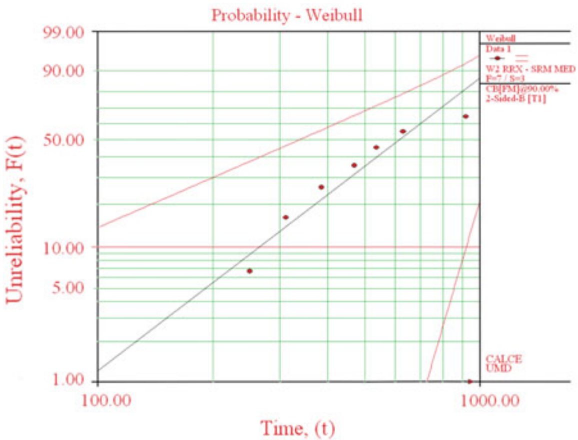


Fig. 23.3 Weibull analysis

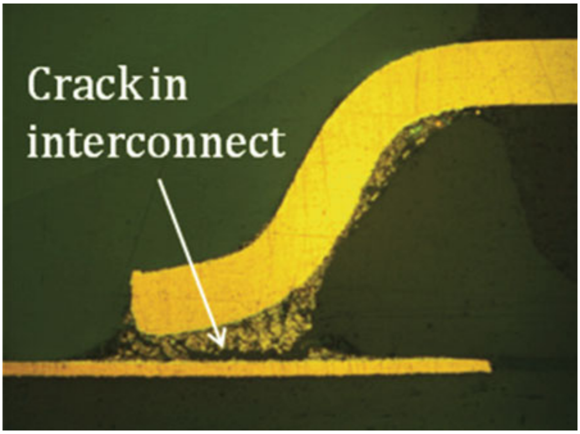


Fig. 23.4 Crack in the solder interface of the TSOP

Figure 23.4 shows a crack in the TSOP package interconnection which resulted in the failure of the flash drive.

23.5 Summary and Conclusions

The electronics industry demands reduced time to market which necessitates shorter product development time. Since experimental testing is time consuming and expensive, simulation based testing of electronic assemblies is gaining wider acceptance. A physics of failure based virtual qualification approach was introduced in this paper. Virtual qualification of the circuit card assembly of a commercially available flash drive was conducted to demonstrate the approach. It was observed that the cycles to failure estimated using the PoF based simulation method were comparable to the accelerated testing results conducted on the flash drive. Analysis of the failed flash drives revealed that the failure was due to fatigue crack propagation at the solder interface in TSOP packages.

References

1. Pecht M, Dasgupta A, Barker D, Leonard CT (1990) The reliability physics approach to failure prediction modeling. Qual Reliab Eng Int 6(4):267–273
2. Pecht M, Dasgupta A (1995) Physics-of-failure: an approach to reliable product development. Proceedings of the integrated reliability workshop, pp 111–117
3. Vandevelde B, Gonzaleza M, Limayea P, Ratcheva P, Beynea E (2007) Thermal cycling reliability of SnAgCu and SnPb solder joints: a comparison for several IC-packages. Microelectron Reliab 47(2–3):259–265
4. Lall P, Gupte S, Choudhary P, Suhling J (2007) Solder joint reliability in electronics under shock and vibration using explicit finite-element submodeling. IEEE Trans Electron Packag Manuf 30(1):74–83

5. Osterman M, Stadterman T (1999) Failure assessment software for circuit card assemblies. Annual proceedings of the reliability and maintainability symposium, 269–276
6. Osterman M, Dasgupta A, Han B (2006) A strain range based model for life assessment of Pb-free SAC solder interconnects. Proceedings of the 56th electronic component technology conference, pp 884–890
7. Cunningham J, Valentin R, Hillman C, Dasgupta A, Osterman M (2001) A demonstration of virtual qualification for the design of electronic hardware. Proceedings of ESTECH 2001, IEST
8. Osterman M, Dasgupta A (2002) Simulation improved testing of electronic hardware. Proceedings of the 48th ATM
9. Osterman M, Pecht M (2007) Strain range fatigue life assessment of lead-free solder interconnects subject to temperature cycle loading. Soldering Surf Mt Technol 19(2):12–17
10. George E, Das D, Osterman M, Pecht M (2009) Physics of failure based virtual testing of communication hardware. ASME International mechanical engineering congress and exposition, pp 191–197
11. Qi H, Chris W, Osterman M, Pecht M (2004) Failure analysis and virtual qualification of PBGA under multiple environmental loading. Proceedings of the 54th electronic component technology conference, vol 1, pp 413–420
12. Engelmaier W (1983) Fatigue life of leadless chip carrier solder joints during power cycling. IEEE Trans Compon Hybrids Manuf Technol 6(3):232–237
13. Engelmaier W (1993) Generic reliability figures of merit design tools for surface mount solder attachments. IEEE Trans Compon Hybrids Manuf Technol 16(1):103–112

Chapter 24
Estimating the Loading Condition of a Diesel Engine Using Instantaneous Angular Speed Analysis

T. R. Lin, A. C. C. Tan, L. Ma and J. Mathew

Abstract Continuous monitoring of diesel engine performance is critical for early detection of fault developments in the engine before they materialize and become a functional failure. Instantaneous crank angular speed (IAS) analysis is one of a few non-intrusive condition monitoring techniques that can be utilized for such tasks. In this experimental study, IAS analysis was employed to estimate the loading condition of a 4-stroke 4-cylinder diesel engine in a laboratory condition. It was shown that IAS analysis can provide useful information about engine speed variation caused by the changing piston momentum and crankshaft acceleration during the engine combustion process. It was also found that the major order component of the IAS spectrum directly associated with the engine firing frequency (at twice the mean shaft revolution speed) can be utilized to estimate the engine loading condition regardless of whether the engine is operating at normal running conditions or in a simulated faulty injector case. The amplitude of this order component follows a clear exponential curve as the loading condition changes. A mathematical relationship was established for the estimation of the engine power output based on the amplitude of the major order component of the measured IAS spectrum.

T. R. Lin (✉) · L. Ma · J. Mathew
CRC for Infrastructure and Engineering Asset Management, Brisbane, QLD 4001, Australia
e-mail: trlin@qut.edu.au

L. Ma
e-mail: l.ma@qut.edu.au

J. Mathew
e-mail: j.mathew@qut.edu.au

T. R. Lin · A. C. C. Tan · L. Ma
School of Engineering Systems, Queensland University of Technology, Brisbane, QLD 4001, Australia
e-mail: a.tan@qut.edu.au

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_24,

24.1 Introduction

Diesel engines are one of the most critical classes of machinery in industry today. Unpredicted failures of diesel engines can cause dire consequences. It is thus essential to mitigate unpredicted engine failures to ensure continuing operation and optimum engine performance. Engine failures could be caused either by combustion related subsystems, such as the fuel injection system, the cylinder/piston system, the inlet and outlet valve system, or by non-combustion related subsystems including all auxiliary devices such as turbochargers, gears, bearings, and electronic control units. Faults developed in combustion related subsystems will directly affect the combustion process and engine performance. Engine misfire, knocking, insufficient power output, poor fuel efficiency, excessive exhaust smoke, excessive noise, and vibration are some of the most typical fault symptoms of a diesel engine. To minimize or to prevent the occurrence of unpredicted engine failures, the operating state, and health of a diesel engine needs to be monitored continually so that a fault symptom and its cause could be diagnosed and dealt with at the early stage before it becomes a functional failure.

Instantaneous crank angular speed and in-cylinder pressure methods are two commonly employed condition monitoring techniques for engine performance monitoring and combustion related fault detections. In-cylinder pressure technique can provide a direct indication of engine performance and the state of the engine combustion process. However, applications of the technique are restrained by the intrusive nature of the method. Furthermore, in-cylinder pressure measurement is a localized technique which can only detect or monitor the combustion process of the cylinder where a pressure sensor is mounted. Prompted by the rapid development of data acquisition hardware and signal processing techniques over the last decade or two, a non intrusive instantaneous angular speed (IAS) measurement is gaining wide applications for condition monitoring and fault detection of rotating machinery. For instance, IAS technique was successfully employed for torsional vibration analysis [1, 2], condition monitoring of gear transmission [2–5], condition monitoring and fault detection of diesel engines [6–9], condition monitoring of electric motors [10] and roller bearing fault detection [11]. Instantaneous angular speed (IAS) technique can also be employed for monitoring engine performance as the variation of crank angular speed is directly correlated to the total gas pressure torque produced by the engine combustion process.

In the application of IAS for fault detection and diagnosis of diesel engines, Yang et al. [6] presented a simple two-degree of freedom dynamic system for the simulation of instantaneous angular speed fluctuation of a four-stroke four-cylinder diesel engine. They found that the instantaneous angular speed is largely affected by the tangential force induced by the gas pressure and the vertical imbalance inertial force induced by piston acceleration/deceleration of all the cylinders. They further illustrated that the fluctuation of IAS can be utilized for detecting engine combustion related faults such as those affecting the gas pressure in the cylinders, i.e., fuel or exhaust valve leakages. Charles et al. [7] extended the IAS technique further for condition monitoring and fault diagnosis of large diesel engines with a high number of cylinders. By presenting the IAS waveform in a polar coordinate system, they demonstrated that IAS waveform can be utilized to detect and identify the faulty (misfiring) cylinder of two relatively large multi-cylinder diesel engines (16 and 20 cylinders respectively). Taraza et al. [8] investigated the amplitude change of order components of IAS waveforms of two diesel engines and correlated the amplitude of the lowest major harmonic order (order 2 for a 4-cylinder engine and order 3 for a six-cylinder engine) of IAS spectra to that of the gas pressure torque produced by the engine combustion. They also illustrated that phases of the three lowest order components of the IAS spectrum could be utilized to identify the faulty cylinder of a diesel engine. Douglas et al. [9] applied both acoustic emission and instantaneous angular speed techniques for on-line power estimation of two large marine diesel engines. They found that the calculated standard deviation of IAS waveforms in each engine cycle changes accordingly to the loading condition of the two marine engines under test. The change patterns also agreed well with those of the measured acoustic emission (AE) root mean squared (RMS) energy per engine cycle at various loading conditions.

Recently, Li et al. [12] presented a comprehensive discussion for optimizing IAS measurements and provided a detailed IAS error analysis. Gu et al. [13] developed a theoretical model for noise reduction of IAS signals by employing a combined FFT and Hilbert transform. They then implemented the method to reduce the noise in IAS data acquisition and its influence on the IAS estimation and fault diagnosis of a rotor-shaft test rig.

The two most common approaches for instantaneous angular speed measurements are time/counter based method and analogue to digital conversion (ADC) based method [12]. ADC-based methods, which use only standard data acquisition equipment, are much easier to implement in practical situations than time/counter based methods. The principle of ADC based methods is to acquire analogue encoder signals from the rotating shaft where the encoder is attached onto and then converts the signal into digital data at a fixed sampling frequency. Thus, the IAS data can be extracted directly from raw encoder signals and the method is less likely to be affected by ambient noise. An ADC based IAS method is employed in this study for the diesel engine performance monitoring and is extended further for diesel engine fault detection which will be reported separately.

The contents presented in this paper are arranged as follows: Sect. 24.2 presents a description of the test rig and the experimental setup for data acquisition and condition monitoring of a diesel engine. An order analysis of IAS waveforms extracted from raw encoder signals are presented in Sect. 24.3. Discussion and insightful comments of the results obtained from this investigation are also offered in the section. Section 24.4 summarizes the main findings from this study.

24.2 The Diesel Engine Test Rig and Experiments

Simulating common diesel engine faults in a controlled manner under laboratory conditions, and characterizing the signal patterns of simulated faults at various loading conditions can provide instantaneous and accurate fault diagnosis of diesel engines since faults in a diesel engine do not normally occur in a short period of time. To this end, a GEP18-4 Olympian diesel engine generator set as shown in Fig. 24.1a was used in the fault simulation and experimental investigation presented in this paper. The specification of the engine is described in Table 24.1. The diesel engine generates a nominal power output of about 15 kW. A three-phase 15 kW industrial fan heater as showed in Fig. 24.1b was used to absorb the power output generated by the diesel engine generator set in the experiment. The fan

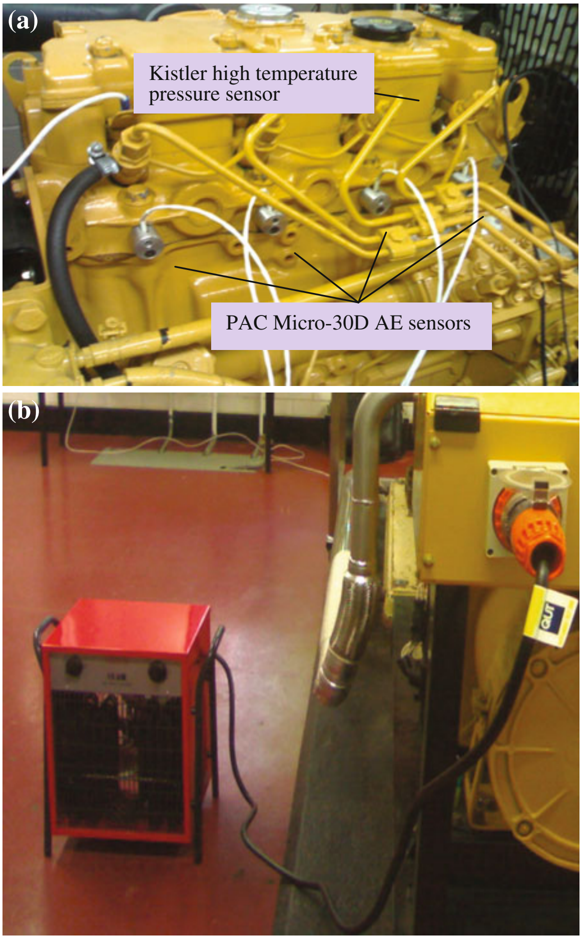


Fig. 24.1 Graphical illustration of the diesel engine test rig and sensors; the test rig (**a**) and the industry fan heater (**b**)

heater which has three heat settings was adjusted for various engine loadings during the experiment.

The sensors used in the simulation study included four resonance type (PAC Micro-30D) acoustic emission sensors, a (Kistler) high temperature pressure sensor and an (PCB) ICP piezoelectric accelerometer as shown in Fig. 24.1. A combined optical encoder and top dead centre (TDC) recording unit taken out from the electronic distributor of a Nissan Bluebird car was also installed onto the crankshaft next to the flywheel for the measurement of instantaneous angular speed and torsional vibration. The encoder has 360 circumferential evenly spaced slits (1° resolution) and it was assumed that the encoder was ideally manufactured with negligible manufacturing errors. The circumferential space between two sequential slits of the encoder is about the same as the span of each slit. The flywheel on the crankshaft reduced engine speed fluctuations, particularly at the unloaded conditions. The other sensors installed on the diesel engine test rig were for comparative studies of their effectiveness in condition monitoring and fault detection under various simulated faults.

A multi-channel National Instrument data acquisition card (DAQ Card-6062E) with a sampling frequency of up to 500 kHz, coupled with a LabVIEW software installed on a laptop computer were used for the data acquisition of encoder signals in the experiment. The sampling frequency in the measurement was set at 100 kHz, which was pre-determined according to the following equation to minimize the IAS measurement errors [13]

$$f_s > 4\left[nf_{shaft} + \left(n_h f_{shaft} + n\Delta f\right)\right], \tag{24.1}$$

where f_s is the sampling frequency, n is the number of circumferential evenly spaced slits of the encoder, n_h is the highest order of IAS components of concern, f_{shaft} is the shaft rotating frequency and Δf is the estimated shaft speed variation.

Table 24.1 Diesel engine specification

Perkins 404C-22 engine data	
Number of cylinders	4
Arrangement	In-line
Running speed	1,500 rpm
Bore	84.0 mm
Stroke	100.0 mm
Displacement	2.216 L
Compression ratio	23.3:1
Firing order	1-3-4-2
Injection timing (estimated)	15° (± 1°) before TDC
Exhaust valve open (measured)	143° after TDC
Exhaust valve close (measured)	370° after TDC
Inlet valve open (measured)	354° after TDC
Inlet valve close (measured)	584° after TDC

The length of each data record was 5 s which encompasses about 124 $\sim$ 130 shaft revolutions in each data file depending on engine loading conditions. Ten (10) data files were recorded for each engine loading at the steady state condition. These data files were used for offline averaging in the order domain during the post processing of the IAS data to minimize the effect of amplitude variation on the signal characterization of the diesel engine.

24.3 Instantaneous Angular Speed Analysis

The measured encoder signal (within a small time window) of the unloaded diesel engine at the normal running condition is shown in Fig. 24.2. The measured encoder signal was used to construct the instantaneous angular speed (IAS) waveform in the study. From the measured encoder signal, the instantaneous angular speed was calculated approximately by

$$\omega(t) = \frac{\partial \varphi(t)}{\partial t} \cong \frac{\varphi_j - \varphi_{j-1}}{\Delta t_j} \qquad (\text{rad/s}), \tag{24.2}$$

where φ_j and φ_{j-1} are the crank angles of two sequential slits of the encoder, and Δt_j is the time difference between the two slits.

When examining the amplitude and the data point distribution of encoder pulses shown in Fig. 24.2, it was found that the distribution of data point locations of each encoder pulse was slightly different from each other. For a more accurate

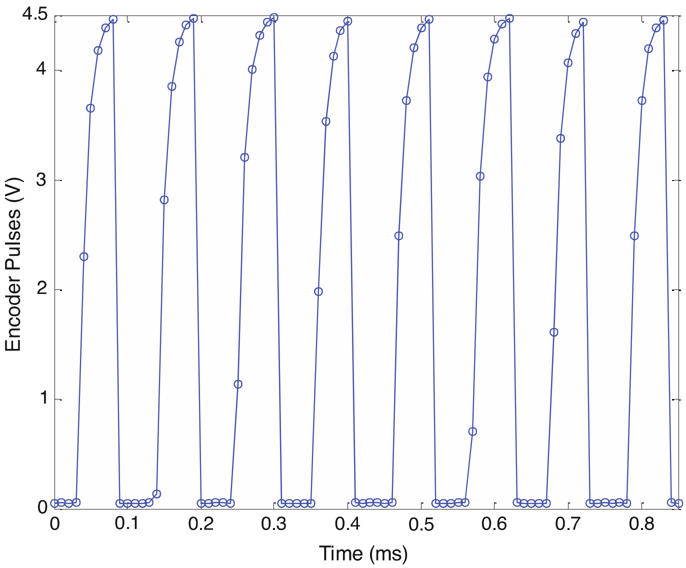


Fig. 24.2 Raw encoder signal at the engine unloaded condition

estimation of the instantaneous crank angular speed, a simple algorithm was implemented in the post processing of the encoder signal. In this implementation, the first data point on the rising edge of each encoder pulse which passes through a pre-defined amplitude threshold was identified first. A linear interpolation algorithm was then employed for a more accurate estimation of the time when a pulse passed the threshold value. An amplitude threshold value of 4 V was chosen after examining the data point distribution in Fig. 24.2 to minimize the time estimation error in the linear interpolation algorithm. Procedures of the linear interpolation algorithm in processing the encoder signal will be described in detail separately.

24.3.1 IAS at Normal Engine Running Conditions

The IAS waveform of the diesel engine at the unloaded condition calculated using the linear interpolation algorithm and Eq. (24.2) is shown in Fig. 24.3. The power spectrum of the IAS waveform displayed in the order domain is shown in Fig. 24.4. The major order components of the IAS spectrum can be correlated to major mechanical events of the diesel engine. For instance, the first major dominant order component of the IAS spectrum is attributed to the shaft rotating speed, the second order component corresponds to the engine firing frequency of the 4-stroke 4-cylinder diesel engine, while the fourth order component corresponds to the four top (bottom) dead centres of the diesel engine per shaft revolution. The identification of major order components of the IAS spectrum to the mechanical events of the diesel engine is not the objective of this work. Instead, the study aims to reveal and correlate the changing pattern of the major order component of the IAS spectrum to the power output (loading) of the diesel engine.

24.3.2 Standard Deviation Per Engine Cycle of the IAS Waveform at Various Loading Conditions

Engine power output estimation was studied by Douglas et al. [9] using an acoustic emission technique and standard deviation per engine cycle of IAS waveforms for two large marine diesel engines. In their work, they studied the change pattern of standard deviation per engine cycle of IAS waveforms at two loading conditions for one of the two test engines and three loading conditions for the other engine. They observed that the changing pattern of standard deviation per engine cycle of IAS waveforms is consistence with the changing power output of the diesel engine at the loading conditions under study. A major drawback of their work is that it did not provide a quantification analysis to establish an explicit mathematical relationship to describe the power output change by the amplitude change of the standard deviation.

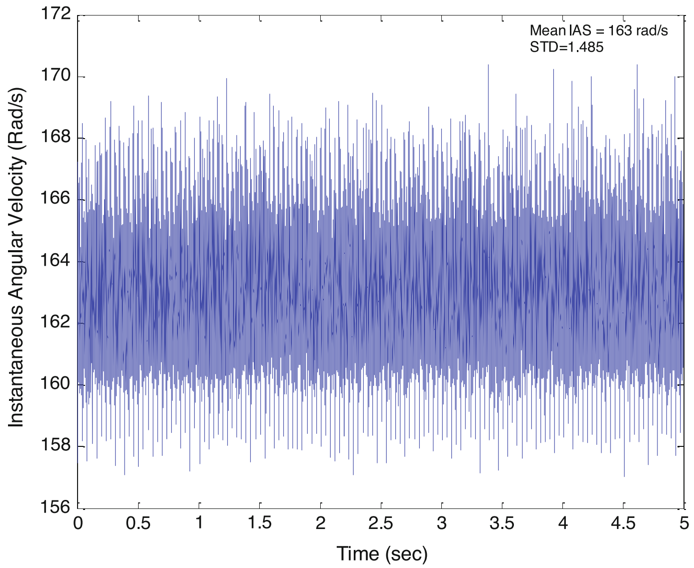


Fig. 24.3 Time waveform of the calculated instantaneous angular speed at an unloaded condition

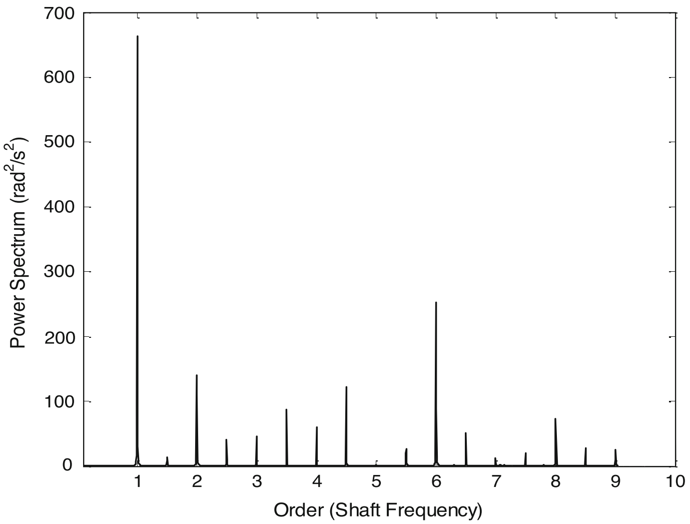


Fig. 24.4 Power spectrum of the IAS waveform at an unload condition

In this study, the standard deviation per engine cycle of IAS waveforms of the diesel engine under four loading conditions, namely, unload, one third load, two third load, and full load, is calculated according to [9]:

$$\sigma_i = \sqrt{\frac{1}{n-1}\sum\nolimits_{j=1}^{n}\left(\omega_j - \overline{\omega}_i\right)}, \tag{24.3}$$

where n is the length of the IAS data of the ith engine cycle, ω_j is the IAS at the time instance t_j and $\overline{\omega}_i$ is the mean IAS for the ith engine cycle.

The results are shown in Fig. 24.5. It is observed that the amplitude of the standard deviation per engine cycle of the IAS waveform decreases as the engine loading condition increases for the first three loading conditions. However, the trend does not continue for the last loading condition when the engine is running at full load. The result indicates that the standard deviation method may have limited applications in practical cases, for instance, to be used for engine loading estimation at low engine loading conditions. To overcome such limitation, an alternative method is developed in this study for diesel engine output power estimation, which is elaborated in the next section.

24.3.3 Correlation Between the Major Order Component of an IAS Spectrum and the Engine Loading Condition

In this section, the correlation relationship between the amplitude of the major order component of the IAS spectrum and the loading condition of the diesel engine is established. The major order component of the IAS spectrum used in the

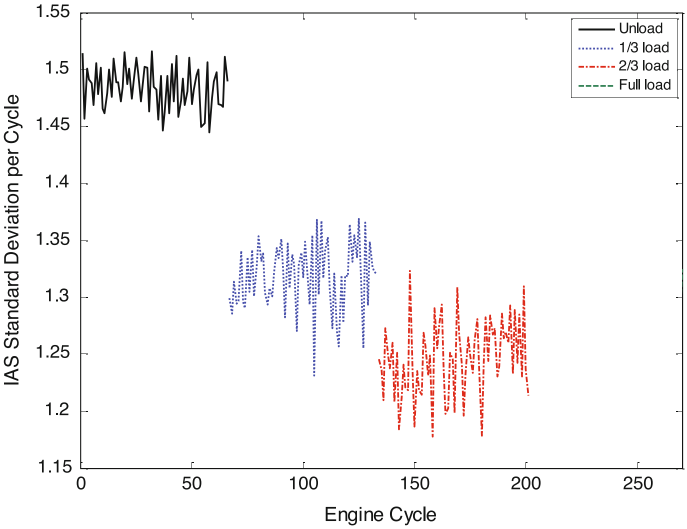


Fig. 24.5 Standard deviation per engine cycle of IAS signals at various engine loading conditions

study is the order component directly associated with the engine firing frequency. To minimize the effect of the engine speed and amplitude variation to the analysis, the IAS spectrum was averaged over the ten encoder files in the order domain for each engine loading condition, namely, unloaded (0 kW), one-third load (5 kW), two-third load (10 kW), and full load (15 kW). The major order component corresponding to the engine firing frequency (combustion) of the IAS spectrum at various loading conditions is shown in Fig. 24.6 for comparison. It is observed that the amplitude of the order component follows a clear exponential curve. This exponential curve can be described by the following equation

$$\mathbf{A} = A_0 \mathrm{e}^{(\alpha * \mathbf{N})}, \quad (\mathrm{rad/s})^2, \tag{24.4}$$

where $\mathbf{A}$ is the amplitude vector of the power spectrum component at various engine loading conditions, A_0 is the amplitude of the IAS power spectrum component at the unload condition, $\mathbf{N}$ is the vector indicating the loading conditions of the engine, which takes the value of 0 for unload condition and 1 for full load condition. $\mathrm{N} = \lfloor 0 \quad 1/3 \quad 2/3 \quad 1 \rfloor^T$ for the case presented in this study where T indicates a vector transpose.

The coefficient α in Eq. (24.4) can be determined by a least square curve fit as

$$\alpha = \frac{\mathrm{N}^T \ln\left(\frac{\mathrm{A}}{A_0}\right)}{\mathrm{N}^T \mathrm{N}}, \tag{24.5}$$

where $\alpha = 1.6873$ was found for the case shown in Fig. 24.6.

Once the relationship between the engine loading condition and the amplitude of the order component is established and the amplitude of the order component at

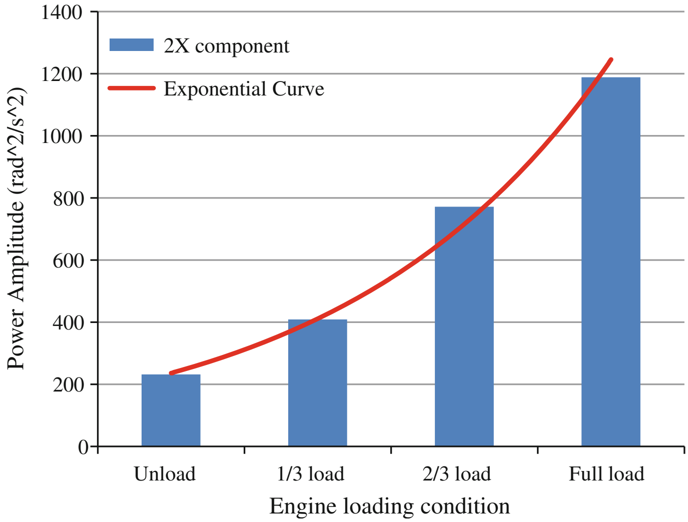


Fig. 24.6 Power amplitude of the order component corresponding to the engine firing frequency

the unload condition is known, the power output of the diesel engine at any particular instance can be estimated by the amplitude (A_j) of the order component of the measured encoder signal at the instance as

$$N_j = \frac{1}{\alpha} \ln \frac{A_j}{A_0}, \tag{24.6}$$

where N_j is the normalized engine power output (with respect to the nominal engine output power) at the t_j time instance.

The trend of amplitude change of this order component (correlates to the engine firing frequency) can be explained by the effect of changing engine loading condition on the engine combustion process and the gas pressure torque. Figure 24.7 shows the averaged in-cylinder pressure of cylinder 1 (the data was averaged over 180 engine cycles using a time waveform event driving synchronizing averaging technique in which the averaging process was correlated and triggered by the major mechanical events of the diesel engine to overcome the difficulty of averaging quasi-periodic signals in the time domain) at the four engine loading conditions. It is illustrated that the increase engine loading not only increased the peak amplitude of the engine combustion pressure (slightly) but also enlarged the area under the in-cylinder pressure curve. This indicates an increase gas pressure torque during the engine combustion when the loading increased, and thus led to the increased amplitude for the order component of the IAS spectrum correlating to the engine (combustion) firing frequency.

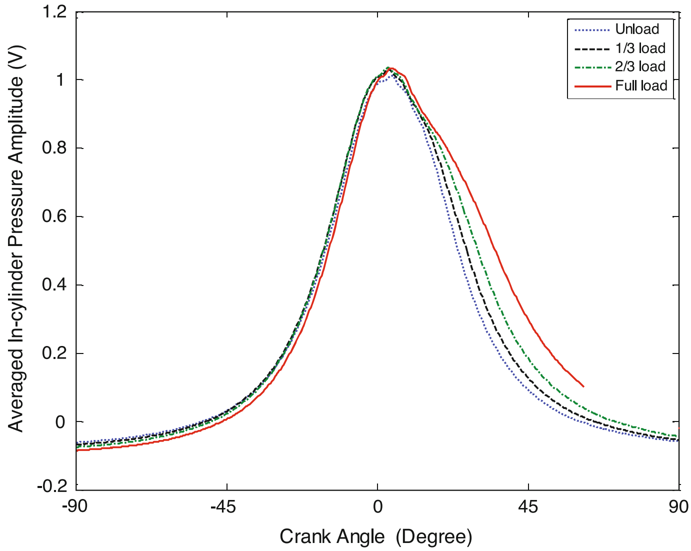


Fig. 24.7 Comparison of the averaged in-cylinder pressure of cylinder 1 at various loading conditions

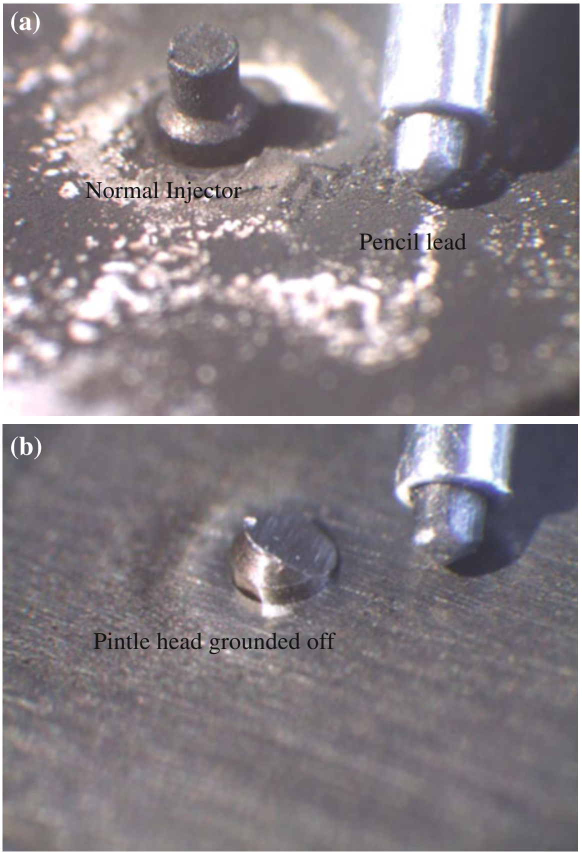


Fig. 24.8 The simulated injector fault; normal injector head (**a**); pintle head partly grounded off (**b**)

24.3.4 The Order Component of the IAS Spectrum at a Simulated Injector Fault Condition

To examine whether the pattern shown in Fig. 24.6 can also be found at other engine operating conditions such as engine running with a faulty injector, one was implanted into cylinder 1 of the engine test rig in the fault simulation experiment. The faulty injector is shown in Fig. 24.8 where a part of the pintle head of the injector was grounded off to simulate a defect of fuel spreading pattern during the fuel injection in the cylinder. It was found from the in-cylinder pressure measurement that the simulated injector fault (representing an incipient injector fault) had only a very small effect on the engine combustion pressure. It was also found that the IAS analysis at various engine loading conditions did not provide useful conclusive information for the detection of this simulated injector fault. This part of the analysis is not presented here to limit the scope of this paper. Examinations of amplitude change of the same order component of the IAS spectrum as showed

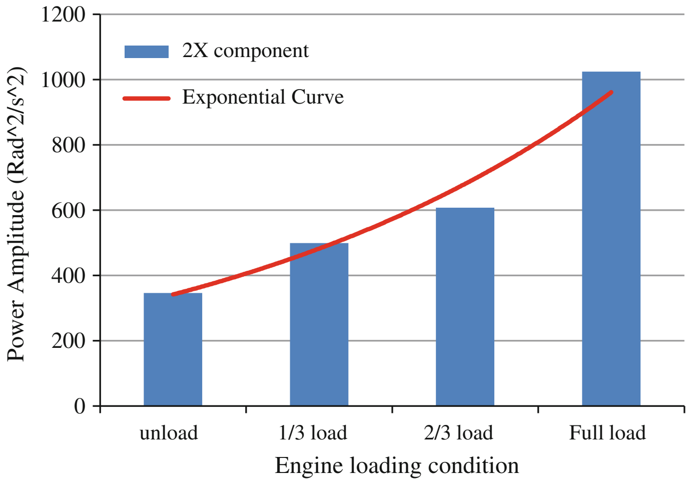


Fig. 24.9 Power amplitude of the order component corresponding to the cylinder firing frequency at the faulty injector case

in Fig. 24.9 revealed that the change in amplitude of the order component also followed an exponential curve in the simulated injector fault case. This curve can be described by an exponential relationship similar to Eq. (24.4) with a slightly modified coefficient α' and the unload base amplitude A'_0 as

$$\mathbf{A}' = A'_0 e^{(\alpha' * \mathbf{N})}. \tag{24.7}$$

24.4 Conclusion

In this paper, an instantaneous angular speed analysis technique is presented for the estimation of engine power output and condition monitoring of diesel engines. It is shown that IAS analysis can provide useful information about engine speed variation caused by changing piston momentum and crankshaft acceleration and can also be used for engine power output estimation in both normal running and faulty injector conditions. It is found that the method of standard deviation per engine cycle of an IAS waveform [9] can only be employed to estimate the diesel engine power output at low loading conditions.

The low frequency order components of the IAS spectrum of a diesel engine can be traced back to major mechanical events of the diesel engine. In particular, it is found in this study that the changing amplitude of the order component of the IAS spectrum corresponding to the engine firing frequency follows a clear exponential curve to the changing engine loading condition. This trend can be explained by the fact that the amplitude of this order component is directly associated with the engine combustion process and the gas pressure torque whose amplitude also increases with increased engine loading conditions. A mathematical relationship was established in the study for the in situ estimation of engine power output based on changing amplitude of the order component of the IAS spectrum corresponding to the engine firing frequency. The revelation of the correlation relationship between this order component and the engine loading condition as well as the mathematical model established in this work, can be employed for engine loading monitoring in practical applications.

Acknowledgments This paper was developed within the CRC for Infrastructure and Engineering Asset Management, established and supported under the Australian Government's Cooperative Research Centres Programme. The authors gratefully acknowledge the financial support provided by the CRC.

References

1. Feldman M, Seibold S (1999) Damage diagnosis of rotors: application of Hilbert transform and multihypothesis testing. J Vib Control 5:421–442
2. Remond D (1998) Practical performances of high-speed measurement of gear transmission error or torsional vibrations with optical encoders. Meas Sci Technol 9:347–353
3. Stander CJ, Heyns PS (2005) Instantaneous angular speed monitoring of gearboxes under non-cyclic stationary load conditions. Mech Syst Signal Process 19(4):817–835
4. Du S, Randall RB (1998) Encoder error analysis in gear transmission error measurement. Proc Inst Mech Eng Part C: J Mech Eng Sci 212:277–285
5. Sweeney PJ, Randall RB (1996) Gear transmission error measurement using phase demodulation. Proc Inst Mech Eng Part C: J Mech Eng Sci 210(3):201–213
6. Yang J, Pu L, Wang Z, Zhou Y, Yan X (2001) Fault detection in a diesel engine by analysing the instantaneous angular speed. Mech Syst Signal Process 15(3):549–564
7. Charles P, Sinha JK, Gu F, Lidstone L, Ball AD (2009) Detecting the crankshaft torsional vibration of diesel engines for combustion related diagnosis. J Sound Vib 321:1171–1185
8. Taraza D, Henein NA, Bryzik W (2001) The frequency analysis of the crankshaft's speed variation: a reliable tool for diesel engine diagnosis. Trans ASME J Eng Gas Turbines Power 123:428–432
9. Douglas RM, Steel JA, Reuben RL, Fog TL (2006) On-line power estimation of large diesel engine using acoustic emission and instantaneous crankshaft angular velocity. Proc Inst Mech Eng Int J Eng Res 7:399–410
10. Sasi AYB, Gu F, Payne B, Ball A (2004) Instantaneous angular speed monitoring of electric motors. J Qual Maintenance Eng 10(2):123–135
11. Renaudin L, Bonnardot F, Musy O, Doray JB, Remond D (1998) Natural roller bearing fault detection by angular measurement of true instantaneous angular speed. Meas Sci Technol 24:1998–2011
12. Li Y, Gu F, Harris G, Ball A, Bennett N, Travis K (2005) The measurement of instantaneous angular speed. Mech Syst Signal Process 19:786–805
13. Gu F, Yesilyurt I, Li Y, Harris G, Ball A (2006) An investigation of the effects of measurement noise in the use of instantaneous angular speed for machine diagnosis. Mech Syst Signal Process 20:1444–1460

Chapter 25
Life and Reliability Prediction of the Multi-Stress Accelerated Life Testing Based on Grey Support Vector Machines

F. Sun, X. Li and T. Jiang

Abstract There are many difficulties in statistical analysis of multi-stress accelerated life testing, such as establishing the accelerated model and solving pluralism likelihood equations. With a focus on these difficulties, the Grey-SVM based life and reliability prediction method for multi-stress accelerated life testing is proposed, with the accelerated stress level and the reliability as SVM inputs, and the corresponding Grey AGO processing failure data as outputs. Simulation and case study shows that the method has high prediction accuracy and with less amount of training samples than neural network.

25.1 Introduction

Accelerated Life Testing (ALT) is an effective way to quickly assess the life and reliability of long-life and high-reliable products [1–3]. The key of the assessment of ALT data is establishing the relationships between the life characteristics and the stress level, i.e., so-called accelerated model or accelerated function.

Currently, for the accelerated model study, the single stress accelerated model is relatively mature, such as the Arrhenius model and Eyring model for temperature, and the Inverse Power Model and Exponential Model for voltage. However,

F. Sun (✉) · X. Li · T. Jiang
School of Reliability and System Engineering, Beihang University,
Xueyuan Road, Beijing 100191, China
e-mail: sunfuqiang@dse.buaa.edu.cn

X. Li
e-mail: leexy@buaa.edu.cn

T. Jiang
e-mail: jtm@buaa.edu.cn

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_25,

the influencing environmental stresses of the products are complex in practice. For example, a product may be influenced by temperature, vibration, and humidity simultaneously. The combined effects of these stresses affect the product life. Therefore, the introduction of comprehensive stress in accelerated testing can not only shorten the test time, but also can more accurately simulate the actual environmental conditions, and more reliable results can be obtained [4]. But there are two insurmountable difficulties in the establishment of multi-stress accelerated models. First, the failure mechanism induced by various stress is different and the stress coupling problems exist, which make it difficult to find an appropriate relationship between life and stress; second, even if the accelerated model is determined by empirical or statistical methods, the form will be very complex and difficult to solve.

In addition, the statistical analysis method assuming that the accelerated model for ALT may have a problem in that the actual situation does not fully comply with acceleration model, which would lead to systematic errors. To deal with these difficulties, Wei [5] proposed a life-prediction method of multi-stress ALT based on BP artificial neural network, which avoids establishing the accelerated model and solving pluralism likelihood equations. However, the method uses least square fitting of the failure data to obtain enough training samples. The irreversible fitting losses some information and make the predicted results to have systematic error.

In order to solve the above problems, this paper utilize the Grey theory and Support Vector Machine (SVM) in the reliability assessment of multi-stress ALT and establish a Grey-SVM model to predict the products life and reliability under normal condition. This method does not need to consider accelerated model and parameter estimation, and make improvement for the method in Ref. [5].

25.2 Grey Theory

The grey theory was first proposed by Deng in 1982 [6], which emphasize investigation of the incomplete information and uncertainty. The grey modeling process includes three basic operations: the accumulated generating operation (AGO), the grey modeling and the inverse accumulated generating operation (IAGO). The GM (1, 1) model is the simplest and most commonly used grey model. According to the grey theory, the stepwise ratios $\sigma(k) = x(k-1)/x(k)$ of the n-component sequence x must be contained in the interval $(e^{-2/(n+1)}, e^{2/(n+1)})$. If the stepwise ratios exceed the interval, some data processing, such as shifting, logarithms, and square root processing, needs to be done.

25.3 Theory of the Support Vector Machine for Regression

Support Vector Machine was developed by Vapnik et al. [7, 8], which is a novel learning machine based on statistical learning theory. SVM minimizes the true risk by principle of Structure Risk Minimization (SRM). In 1997, with the introduction of Vapnik's ε-insensitive loss function, SVM was successfully utilized to solve the regression problems. The basic idea of SVM for regression is introduced as follows [9, 10].

Given a set of training data $\{(x_i, y_i), \ldots, (x_l, y_l)\}, i = 1, 2, \ldots, l$, where $x_i \in \mathbf{R}^n$ denote input patterns, $y_i \in \mathbf{R}$ are the targets and l is the total number of training samples. In SVM for regression, the goal is to find a function $f(x)$, i.e., an optimal hyperplane, which has at most ε deviation from the actually obtained targets y_i for all the training data and is as flat as possible. The form of functions is denoted as

$$y = f(x) = \langle \omega, \Phi(x) \rangle + b \quad \text{with } \Phi : R^n \to F, \omega \in F, \tag{25.1}$$

where $\Phi(\cdot)$ is a nonlinear mapping by which the input data x is mapped into a high dimensional space $\boldsymbol{F}$, and $\langle \cdot, \cdot \rangle$ denotes the dot product in space $\boldsymbol{F}$. The unknown variables ω and b are estimated by minimizing the regularized risk function $R_{reg}[f]$

$$R_{\text{reg}}[f] = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{l} \zeta(f(x_i) - y_i), \tag{25.2}$$

$$\zeta(f(x) - y) = \max\{0, |y - f(x)| - \varepsilon\}, \tag{25.3}$$

where C is a constant determining the trade-off between the flatness of the regularized term ($\|\omega\|^2/2$) and the empirical error (the second term) [7, 8]. $\zeta(\cdot)$ is the ε-insensitive loss function, and e is the tube size.

By introducing the non-negative slack variables $\xi_i^{(*)}$, Eq. 25.2 is transformed into the following convex constrained optimization problem:

$$\text{Minimize} \quad \Gamma(\omega, \xi, \xi^*) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{l} \zeta(\xi_i + \xi_i^*), \tag{25.4}$$

$$\text{Subject to} \quad \begin{array}{c} \langle \omega, \Phi(x_i) \rangle + b - y_i \le \varepsilon + \xi_i \\ y_i - \langle \omega, \Phi(x_i) \rangle - b \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, \; i = 1, 2, \cdots, l. \end{array}, \tag{25.5}$$

According to Wolfe's Dual Theorem and the saddle-point condition, the dual optimization problem of the Eq. 25.4 is obtained as the following form:

$$\text{Minimize: } W\left(\alpha_i^{(*)}\right) = \frac{1}{2}\sum_{i,j=1}^{l}\left(\alpha_i^* - \alpha_i\right)\left(\alpha_j^* - \alpha_j\right)\left\langle \Phi(x_i), \Phi\left(x_j\right)\right\rangle$$
$$+\,\varepsilon\sum_{i=1}^{l}\left(\alpha_i^* + \alpha_i\right) - \sum_{i=1}^{l} y_i\left(\alpha_i^* - \alpha_i\right), \tag{25.6}$$

$$\text{Subject to } \quad \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \qquad 0 \le \alpha_i^{(*)} \le C, \quad i = 1, 2, \ldots, l. \tag{25.7}$$

$$\text{With } \omega = \sum_{i=1}^{l}\left(\alpha_i^* - \alpha_i\right)\Phi\left(x_i\right)', \tag{25.8}$$

where $\alpha_i^{(*)}$ are the nonnegative Lagrange multipliers that can be obtained by solving the convex quadratic programming problem stated above. Finally, by exploiting the Karush–Kuhn–Tucker (KKT) conditions, the decision function given by Eq. 25.6 gets the following form:

$$f(x) = \sum_{i=1}^{l}\left(\alpha_i^* - \alpha_i\right)\langle \Phi(x_i), \Phi(x)\rangle + b, \tag{25.9}$$

In order to get the dot product in the feature space simply and avoid the curse of dimensionality, the kernel function $k\left(x_i, x_j\right) = \left\langle \Phi(x_i),\ \Phi\left(x_j\right)\right\rangle$ is introduced, which satisfies the Mercer condition. The most commonly used kernel functions are:

1. The polynomial kernel $k(x, x') = (\langle x, x'\rangle + c)^p,\ p \in N, c \ge 0$
2. The Radial Basis Function (RBF) kernel $k(x, x') = \exp\left(-\|x - x'\|^2 \big/ (2\sigma^2)\right)$
3. The Sigmoid kernel $k(x, x') = \tanh\left(b\langle x, x'\rangle + \theta\right)$.

25.4 Grey-SVM Based Life and Reliability Prediction Method for Multi-Stress ALT

Grey-SVM based life and reliability prediction method for multi-stress accelerated life testing is a non-parametric forecasting method, which not need to establish accelerate model and solve the complex pluralism likelihood equations. With the inerratic failure time data pre-processed by the grey theory as SVM training samples, the SVM prediction model is constructed. Through self-learning, SVM can automatically summarize the law from train data and utilize the law to predict unknown information. The predicted values of life and reliability under normal condition should be obtained by the inverse transform of output information.

This paper aims to non-repairable product, the life prediction of the complete sample constant accelerated life testing under multi-stress is studied. Suppose that there are $k(k > 2)$ stress level determined by two single stress, denoted $S_i(i = 0,\ldots,k)$. Take S_0 represent the normal condition. Given N_i units under the i-th accelerated stress level S_i and denoted the failure time of the j-th unit under i-th stress level as t_{ij}. This paper wants to solve the problem that the life prediction under normal condition S_0 by utilized the failure time data under accelerated stress S_1, S_2, …, S_{k-1}, S_k. Implementation steps are as follows.

First, the failure data under every accelerated stress level are collected through ALT. The reliability corresponding to failure time is calculated by empirical distribution, denoted $R_i\left(t_{ij}\right)$, $i = 1, 2,\ldots,k$, $j = 1, 2,\ldots,N_i$. Thus, the reliability-failure time curve under stress S_i should be gained.

Second, the stepwise ratios test for the failure data under every accelerated stress is conducted, respectively, to determine whether the stepwise ratios to be contained in the accepted interval. If the stepwise ratios exceed the interval, square root processing needs to be done until it is meet the requirement. Consequently, AGO of the Grey processing data were conducted. We can get the AGO failure time t'_{ij} and the corresponding reliability $R'_i(t'_{ij})$

Third, the Grey SVM prediction model is constructed, with the accelerated stresses S_i and the reliability $R'_i(t'_{ij})$ as training inputs, and the corresponding AGO sequence t'_{ij} as outputs.

Finally, the AGO failure data t'_{0j} under normal condition should be predicted by inputting the normal condition S_0 and the setting reliability $R'_0(t'_{0j})$ to Grey SVM model. Then the failure time under normal condition can be obtained by the inverse transform of the predicted values of AGO failure data. We can get the reliability under normal condition by Weibull distribution fitting of the predicted failure time.

The flow diagram of the Grey-SVM based life and reliability prediction method for multi-stress accelerated life testing is shown in Fig. 25.1.

25.5 Case Study

Suppose that the life follows Weibull distribution,

$$F(x) = 1 - \exp\{(t/\eta)^m\}, t \geq 0, \tag{25.10}$$

where, m denotes shape parameter, which is a constant. η denotes characteristic life. If the life of product is mainly influenced by Temperature and Voltage, the accelerated relationship between η and stress is generalized Eyring model, i.e.,

$$\eta = \frac{A}{T} exp\left(\frac{B}{kT}\right) exp\left(V\left(C + \frac{D}{kT}\right)\right), \tag{25.11}$$

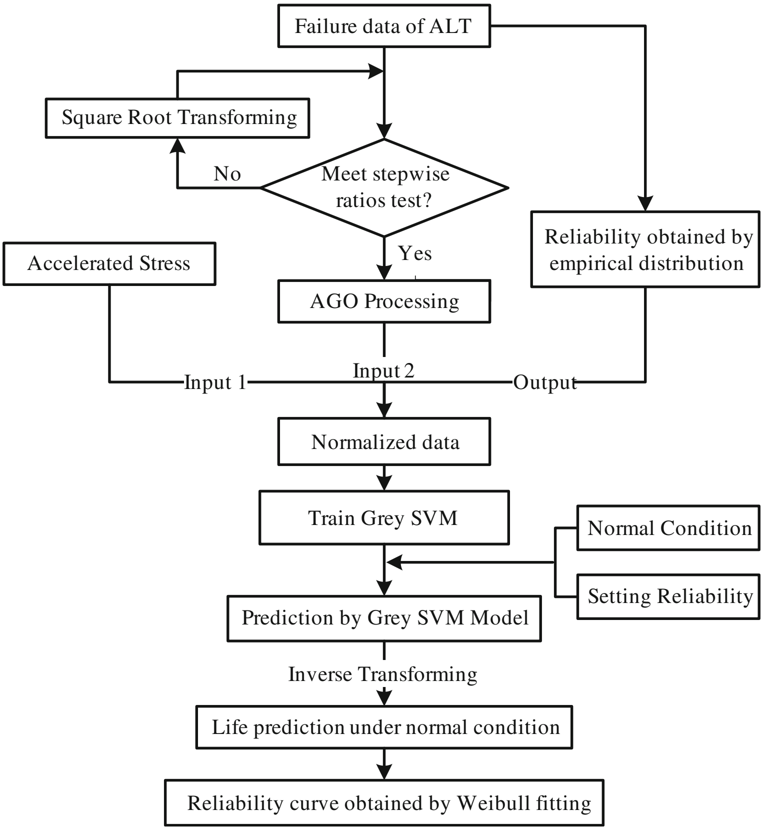


Fig. 25.1 Life prediction flow diagram

where T is the absolute temperature in unit K; V is voltage in unit Volt; A, B, C and D are constant; and $k = 8.6171 \times 10^{-5}$ eV/K is Boltzmann constant. The simulation model is as follows

$$F(x) = 1 - \exp\left\{\left[t \Big/ \left(\frac{A}{T} exp\left(\frac{B}{kT}\right) exp\left(V\left(C + \frac{D}{kT}\right)\right)\right)\right]^m\right\}. \qquad (25.12)$$

The parameters of the simulation model are shown in Table 25.1.

A constant-stress accelerated life testing with 6 stress levels is arranged, under the stress temperature and voltage. The sample size of every stress level is all 10.

Table 25.1 Parameters of simulation model

Parameters	m	A	B	C	D
Values	1.5	15	0.38	−0.48	0.007

Table 25.2 Stress values of simulation model

Stress level	S_0	S_1	S_2	S_3	S_4	S_5	S_6
Temperature/K	310	324	324	339	310	339	355
Voltage/V	12	12	14	14	16	16	16

The setting temperature and voltage are shown in Table 25.2, and S_0 denotes normal condition.

The failure data are obtained by the Monte Carlo simulation, and the reliability corresponding to failure time is calculated by empirical distribution. The reliability- failure time curves are shown in Fig. 25.2.

According to the grey theory, the stepwise ratios of each stress sequence should be contained in the interval $(e^{-2/(10+1)}, e^{2/(10+1)})$. But the stepwise ratios of the sequences cannot meet this requirement. After making square root processing for the sequences, the new sequences are qualified. Consequently, the AGO sequences were obtained, as shown in Figs. 25.3, 25.4.

With the accelerated stresses S_1, S_2, S_3, S_4, S_5, S_6 and the reliability as inputs, the corresponding AGO sequence as outputs, the Grey SVM is trained and the prediction model can be established. By inputting the normal condition and the setting reliability, we can predict AGO failure data under normal condition, as shown in.

Then, prediction of the failure time under normal condition is obtained by the inverse transform of the predicted values of AGO failure data, as shown in Fig. 25.5. We can see that the predicted values and the simulation failure time are

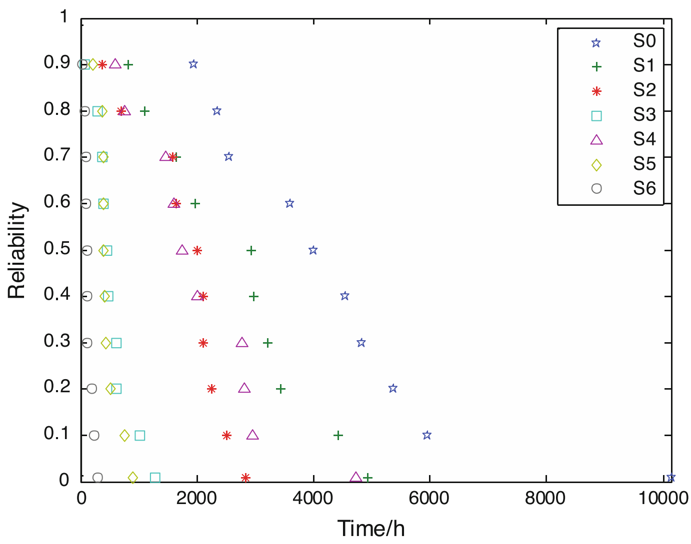


Fig. 25.2 Original failure data

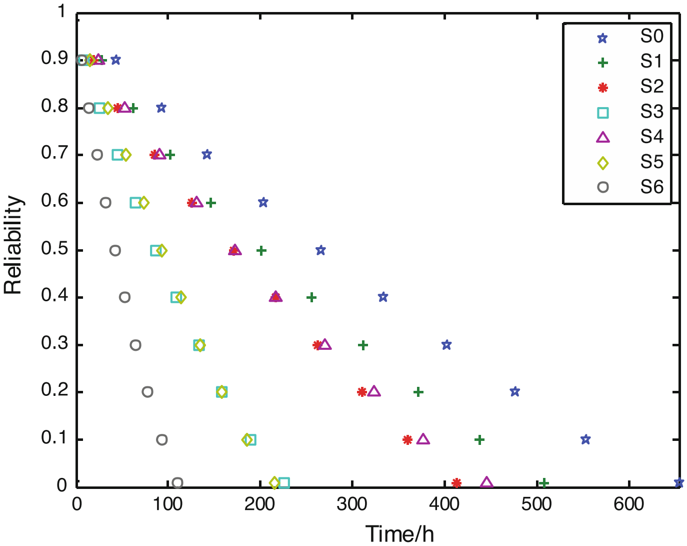


Fig. 25.3 AGO failure data

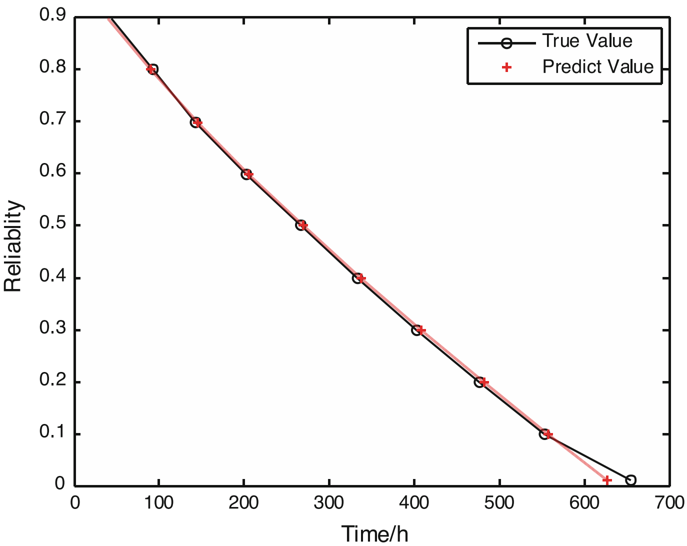


Fig. 25.4 Prediction of AGO failure data

very consistent, but the error is relatively larger on the boundary. Finally, the Weibull distribution fitting of the predicted failure time under normal condition is conducted and the reliability curve is shown in Fig. 25.6. We can obtain all the life information by this curve.

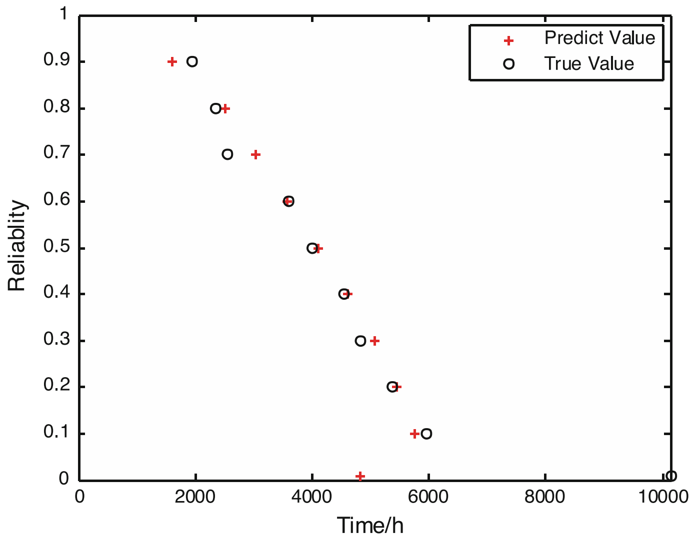


Fig. 25.5 Life prediction under normal condition

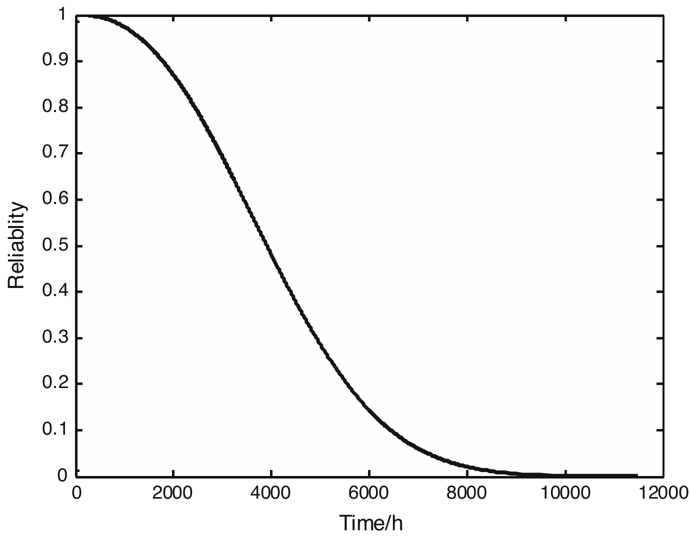


Fig. 25.6 Reliability obtained by Weibull fitting

25.6 Conclusion

Based on combining the advantage of the Grey system theory and SVM, This paper utilizes Grey AGO to process the SVM train data and proposes a new prediction method of multi-stress ALT. This method has following advantages:

(1) not need to know the accelerated model and life distribution, which avoid establishing accelerated model and introducing the systematic error; (2) not need to solve complex pluralism likelihood equations; (3) better engineering applicability for different products or stresses; (4) be suitable for small sample data and facilitate the engineering applications.

However, this method also has some limitations, such as the sample require higher consistency, and so on. These issues need to be further studied and improved.

References

1. Meeker WQ, Escobar LA (1998) Statistical methods for reliability data. Wiley, New York
2. Meeker WQ, Escobar LA, Lu CJ (1998) Accelerated degradation tests: modeling and analysis. Technometrics 40(2):89–99
3. Nelson W (1990) Accelerated testing: statistical models, test plan, and data analyses. Wiley, New York
4. Xiaoyang L, Tongmin J (2007) Review of multiple-stress models in accelerated life testing. Syst Eng Electron 29(5):828–831
5. Wei Z, Xiaoyang L et al (2009) Life-prediction of multi-stress accelerated life testing based on BP algorithm of artificial neutral network. Acta Aeronautica Et Astronautica Sinica 30(9):1691–1696
6. Julong D (2002) Grey prediction and grey decision. Huazhong University of Science and Technology Press, Wuhan, pp 71–75
7. Vapnik VN (1995) The nature of statistical learning theory. Springer, New York
8. Smola AJ, Scholkopf B (2004) A tutorial on support vector regression. Stat Comput 14(3):199–222
9. Xu G, Tian W, Jin Z (2006) An AGO–SVM drift modeling method for a dynamically tuned gyroscope. Meas Sci Technol 17(1):161–167
10. Xu G, Tian W, Qian L (2007) EMD- and SVM-based temperature drift modeling and compensation for a dynamically tuned gyroscope (DTG). Mech Syst Signal Process 21(8):3182–3188

Chapter 26
Reliability Analysis Based on Improved Dynamic Fault Tree

J. Hao, L. Zhang and L. Wei

Abstract Dynamic fault tree analysis has been widely applied to the reliability and safety analysis of various fault tolerant systems. Modern equipment has become larger, more complex and more intelligent; the reliability of this equipment not only impacts the safe operation of personnel and property, but also has important social and economic significance. Based on the improved dynamic fault tree analysis method, this paper analyzes the reliability of oil pump units within a pump station. This paper gives three criteria for ranking the basic events in a Binary Decision Diagram and a comprehensive dynamic fault tree. This method is then put into practice in the reliability analysis of pump units. The result indicates that the improved method used for reliability analysis of oil pump units in a pump station has a simple modeling process, and low computational complexity.

26.1 Introduction

Failure of critical equipment in the oil and gas industry can result in huge personnel and economic losses. Consequently, the need for highly reliable has increasing the complexity of equipment, which has created challenges in reliability analysis. In 1961, bell telephone invented fault tree analysis methods, and used it to predict the stochastic malfunction of missile launchers. At present the method of fault tree analysis has been widely used in electronics, electric power, petrochemical, machinery, transportation and other fields. It can be used to diagnose faults, be used in performing reliability analysis, and realize the optimization of system design. Fault tree analysis is a graphic deductive reasoning method, which can express the logical relationship between the events and consist of familiar with

J. Hao (✉) · L. Zhang · L. Wei
College of Mechanical and Transportation Engineering,
China University of Petroleum, Beijing 102249, China
e-mail: lweimm@yahoo.com.cn

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_26,

system, determining the top event, building the fault tree, quantitative analysis and results analysis. However, the traditional fault tree analysis is a method based on static logic or static fault mechanism, and it's not available for fault-tolerant system and sequence correlation system. So the paper introduces the dynamic fault tree analysis method to solve the problem of reliability of pump units.

26.2 Dynamic Fault Tree Analysis Method

The concept of the dynamic fault tree was first suggested by Dugan [1]. Based on the traditional fault tree, DFT adds some new gates, which reflect the logic of timing sequences relationship of dynamic failure. DFT method combines the advantages of binary decision diagram and Markov transfer chain, introduces new logic gates with dynamic characteristics and sets up a corresponding dynamic fault tree, to provide related references for equipment reliability analysis.

Research on DFT has attracted many scholars' interest in recent years: Antoine B. Rauzy introduced sequence algebras that can be seen as algebras of basic events, representing failures of non- repairable components, and then he proposed a new data structure called Sequence Decision Diagrams [2]. K. Durga Rao and V. Gopika used Monte Carlo simulation method in DFT to deal with probability safety evaluation. Although the calculation time is long, it can avoid the many disadvantages of traditional methods [3]. Chin-Yu Huang improved the decomposition of DFT, and evaluated the reliability and sensitivity of embedded system. They also developed corresponding software, which has very good analysis performance and efficiency [4]. Zhang Honglin gave a DFT method based on isomorphism node, solved the problem of Markov chain state space explosion. Results showed that the method had obvious advantage for use with system structure with strong redundant features [5]. Sun Yuan adopted DFT method to analyze the reliability of automatic control system of subway [6]. With modern equipment becoming larger and complex, the application of DFT will become more and more widely.

Next, the author will think readers have the theoretical basis of the fault tree, mainly introduces new characteristics of dynamic fault tree. Traditional fault tree involves logic gates including "and" gate, "or" gate, etc. In order to express the complexity of modern large-scale fault-tolerant equipment, we need introduce some typical dynamic logic gates:

(1) PAND Gate and SEQ Gate

"PAND" gate is similar to "AND" gate by the logic. It has two input events, and if the top event occurs, the occurrence of the input events must be in some order. As shown in Fig. 26.1, the input events are A and B. If A and B have occurred, and B occurred later than A, the top event can occur. Except this condition, the top event will not occur.

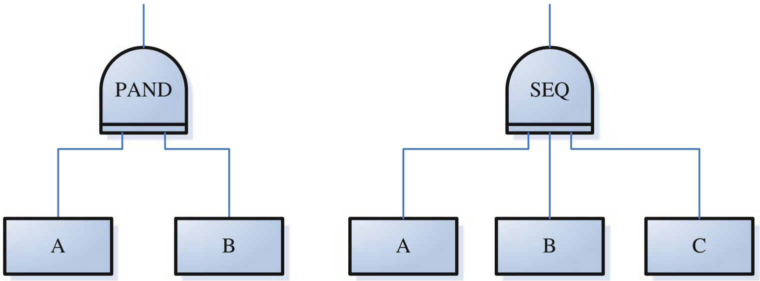


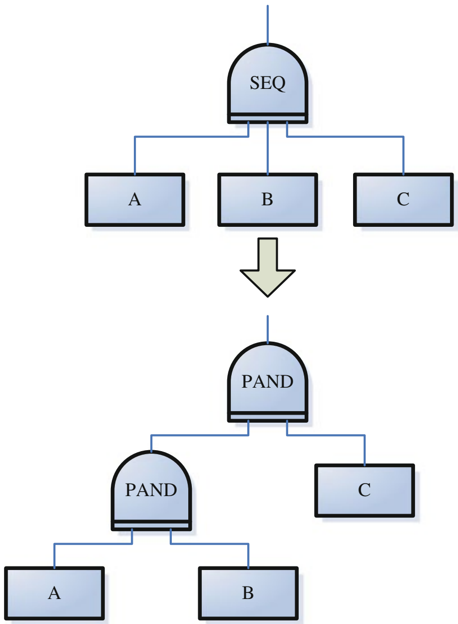
Fig. 26.1 PAND gate and SEQ gate

"SEQ" gate requires input events must occur in accordance with the order from left to right. It's the only way to make the output event occur, otherwise, the output event will not. As shown in Fig. 26.1, the input events can be multiple, and all input events must be the basic event except the far left event. It is not hard to see that PAND gate is the special circumstance of the SEQ gate, and they can signify the time sequence of the system. SEQ gate can be instead by multiple PAND gates, as shown in Fig. 26.2.

(2) FDEP Gate

FDEP gate consists of only a trigger event and several basic events. The failure of some component may lead to other related components failure. Related events must occur when trigger event occur, as shown in Fig. 26.3.

Fig. 26.2 SEQ gate transforms to PAND gate



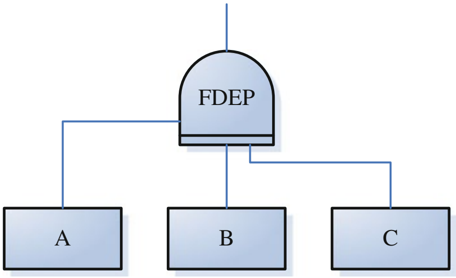


Fig. 26.3 FDEP gate

(3) CSP Gate, WSP Gate and HSP Gate

Cold spare parts refer to the spare parts will not fail before work. The traditional fault tree cannot model the system that has cold spare parts. CSP gate consists of one basic event and more than one optional input events. All the input events are basic events. Basic event are in the work state at the very start, while the optional input events are not and they are the replacements of the basic event. The output event occurs after all input events have occurred, as shown in Fig. 26.4.

WSP and HSP have the failure rate before they enter work status. In some higher reliability systems, people often adopt HSP to switch working state at any time. No matter work or not, HSP have the same failure rate, as shown in Fig. 26.4.

(4) Public Spare Parts Pool [7]

As shown in Fig. 26.5, two spare gates share a public spare parts pool, which was signified by two $S*2$. Because of the specific characteristic of public spare parts pool, which spare parts by which the spare parts door use is unknown. So when one of the spare parts S failed, may lead to any spare part gate failure.

(5) Steps of Dynamic Fault Tree Analysis

Based on traditional method, DFT analysis method is a new reliability analysis method combining with Markov model. As shown in Fig. 26.6, this method first divides the DFT into static and dynamic modules, and then uses Binary Decision Diagram method and Markov Status Transfer Chain method to solve them respectively (Fig. 26.6).

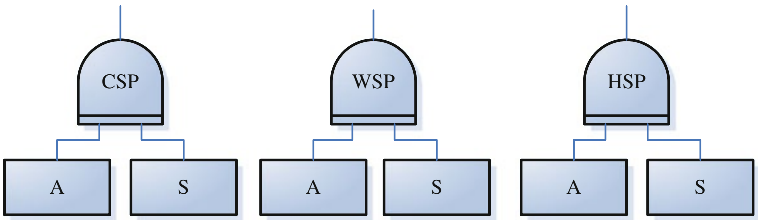
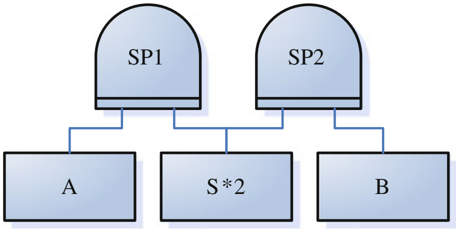


Fig. 26.4 CSP gate, WSP gate and HSP gate

Fig. 26.5 Public spare parts pool



26.3 Binary Decision Diagram and Markov Status Transfer Chain Method

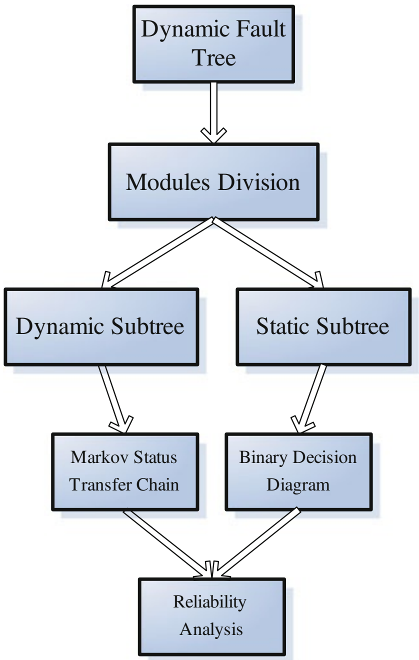
26.3.1 Binary Decision Diagram Method [8]

1. Concept of Binary Decision Diagram

The so-called Binary Decision Diagram is a special rooted tree (V, N), including V as a node set and N an index set. Node set includes two kinds of node:

(1) Leaf Nodes: have a definite value $\{0, 1\}$;
(2) Non-Leaf Nodes: although they have no definite value, the index $(v) \in N$;

Fig. 26.6 Steps of DFT analysis



For any Binary Decision Diagram (V, N), we assume $V = (v_1, v_2 \ldots v_m)$, $N = (1,\ \ 2,\ \ 3,\ \ n)$. Giving a set of Boolean Algebra $x_1, x_2, \ldots, x_n$. According to the corresponding relation, we can make sure that one-to-one correspondence between node and Boolean Algebra, that is index $(v_i) = j \in N, i \in \{1, 2, \ldots, m\}$. So the Boolean Algebra of v_1 is x_j and Binary Decision Diagram can correspond to Boolean Algebra $f = f(x_1, x_2, \ldots x_n)$.

Fault Tree translated into Binary Decision Diagram was first put forward by Rauzy [9]. He used a ite (if-then-else) structure. It refers to: if A is established, then B established, or C as shown in Fig. 26.7. The mathematical expression is (Fig. 26.7):

$$ite(A, B, C) = AB + AC. \tag{26.1}$$

2. Fault Tree transforming into Binary Decision Diagram

There are many ways that transform Fault Tree into Binary Decision Diagram. Here I introduce a method based on recursive thought:

Assuming that the basic event's corresponding Boolean Algebras are $x_1, x_2, \ldots x_n$, and the order of the index is:

$$\mathrm{Index}(x_1) < \mathrm{Index}(x_2) < \ldots\ldots < \mathrm{Index}(x_n).$$

Arbitrary $x_i, x_j \in \{x_1, x_2, \ldots x_n\}$, and defining that:

$$J = ite(x, F1, F2). \tag{26.2}$$

$$H = ite\left(x_j, G1, G2\right). \tag{26.3}$$

The algorithms of the ite structure are:

if $i > j$:

$$J < \mathrm{op} > H = ite(x_i, F1 < \mathrm{op} > H, F2 < \mathrm{op} > H); \tag{26.4}$$

if $i = j$:

$$J < \mathrm{op} > H = ite(x, F1 < \mathrm{op} > G1, F2 < \mathrm{op} > G2). \tag{26.5}$$

$< \mathrm{op} >$ correspond to "or" gate or "and" gate. So there is clearly:

$$1 < \mathrm{op} > H = 1, \tag{26.6}$$

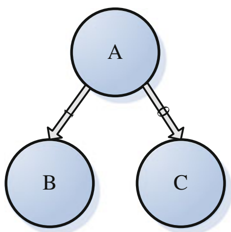


Fig. 26.7 Binary tree structure of $ite(A, B, C)$

$$0 < \mathrm{op} > H = H, \tag{26.7}$$

$$1 < \mathrm{and} > H = H, \tag{26.8}$$

$$0 < \mathrm{and} > H = 0. \tag{26.9}$$

Then take Fig. 26.8 for example, we demonstrate recursive algorithm:

First we assume index $(A1) < \mathrm{index}(A2) < \ldots\ldots < \mathrm{index}(A5)$, according to the layering thought in literature 12, the tree can be divided into five layers. The fifth layer is the bottom. We use recursion method from bottom to top:

The fourth layer:

$$G9 = A4 \cdot A5 = ite(A4, A5, 0), \tag{26.10}$$

$$G10 = A1 \cdot A2 = ite(A1, A2, 0), \tag{26.11}$$

$$G11 = A4 \cdot A5 = ite(A4, A5, 0). \tag{26.12}$$

.

The third layer:

$$G4 = A3 + G9 = ite(A3, 1, ite(A4, A5, 0)), \tag{26.13}$$

$$G5 = G10 + G11 = ite(A1, ite(A2, 1, ite(A4, A5, 0)), ite(A4, A5, 0)), \tag{26.14}$$

$$G6 = A1 \cdot A3 = ite(A1, A3, 0) \tag{26.15}$$

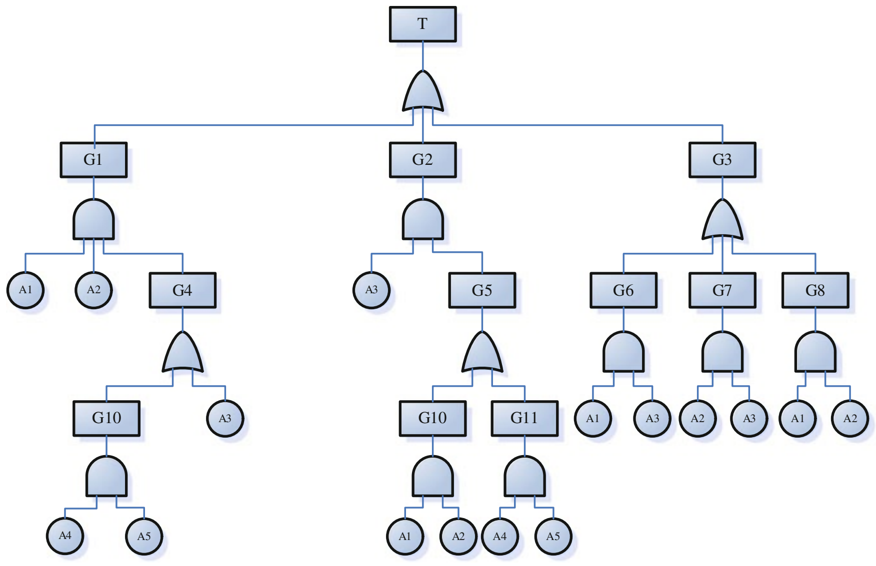


Fig. 26.8 Fault tree example

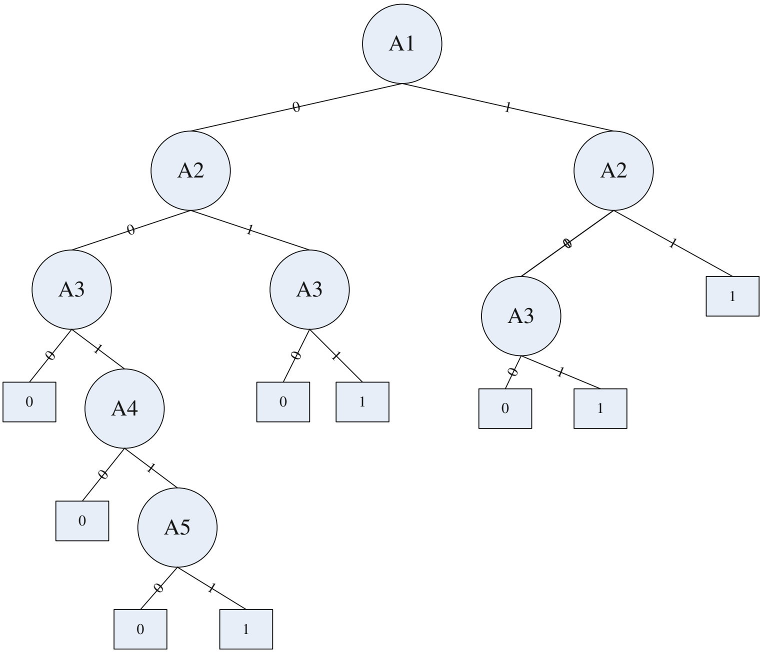


Fig. 26.9 Binary decision diagram of fault tree example

$$G7 = A2 \cdot A3 = ite(A2, A3, 0), \tag{26.16}$$

$$G8 = A1 \cdot A2 = ite(A1, A2, 0). \tag{26.17}$$

The second layer:

$$G1 = A1 \cdot A2 \cdot G4 = ite(A1, ite(A2, ite(A3, 1, ite(A4, A5, 0)), 0)0), \tag{26.18}$$

$$G2 = A3 \cdot G5 = ite \\ (A1, ite(A2, ite(A3, 1, 0), ite(A3, ite(A4, A5, 0), 0)), ite(A3, ite(A4, A5, 0), 0)), \tag{26.19}$$

$$G3 = G6 + G7 + G8 = ite(A1, ite(A2, 1, ite(A3, 1, 0), ite(A2, A3, 0))). \tag{26.20}$$

The first layer:

$$\begin{aligned} T &= G1 + G2 + G3 \\ &= ite(A1, ite(A2, 1, ite(A3, 1, 0)), ite(A2, ite(A3, 1, 0), ite(A3, ite(A4, ite(A5, 1, 0), 0), 0), 0)) \end{aligned} \tag{26.21}$$

So through the expression T, we can get the Binary Decision Diagram of Fault Tree example, as shown in Fig. 26.9.

26.3.2 Markov Status Transfer Chain Method

1. Definition of Markov Status Transfer Chain

Markov process can be described as the result of any moment that only depends on the former moment and has nothing to do with the moment before the former moment. In Markov Chain $x(t)$, if we provide the moment value, the future value cannot be affected by the former status, that is, if $t_{n-1} < t_n$:

$$P[x(t_n) \leq x_n)|x(t), t \leq t_{n-1}] = P[x(t_n) \leq x_n)|x(tn-1)] \tag{26.22}$$

According to the continuity or discreteness of the status and time parameters of Markov process, it can be divided into Markov chain, Markov chain with continuous time and Markov process. Here we mainly discuss the Markov chain with consecutive time and discrete state

2. Markov Transfer Chain of Dynamic Logic Gate
(1) Markov Transfer Chain of PAND and SEQ (Fig. 26.10)
(2) Markov Transfer Chain of FDEP (Fig. 26.11)
(3) Markov Transfer Chain of CSP, WSP and HSP (Fig. 26.12)
(4) Markov Transfer Chain of Public Spare Parts Pool

As shown in Fig. 26.13, when the failure of Public Spare Parts Pool S^*2 may lead to failure of any spare part: SP1 or SP2. Despite the failure of spare part gate is different, the basic lapsed events are same. We call this state "Split State".

Take Fig. 26.14 for example, two HSP gate (P1, P2) share a basic event S, we assume that:

A Failure of P1, P2 or S;
B Failure of P2 or S(S is the substitute for P2);
C Failure of P1 or S(S is the substitute for P1);

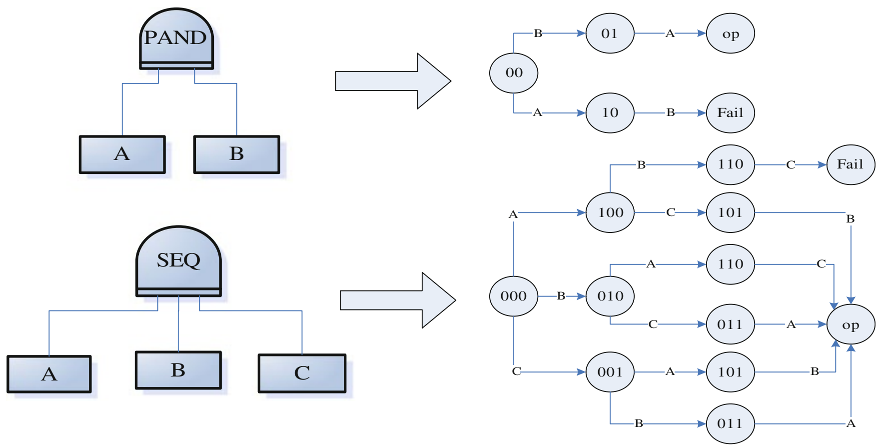


Fig. 26.10 Markov transfer chain of PAND and SEQ

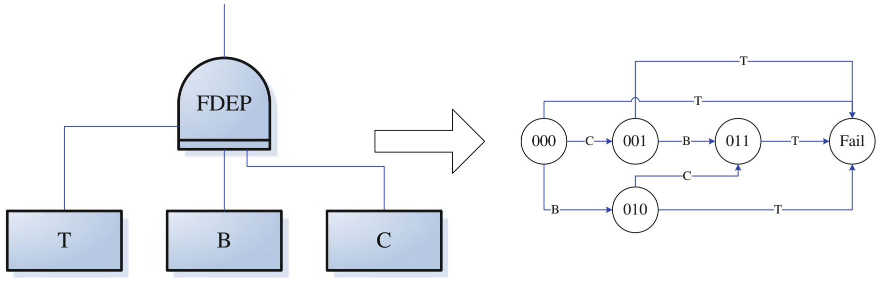


Fig. 26.11 Markov transfer chain of FDEP

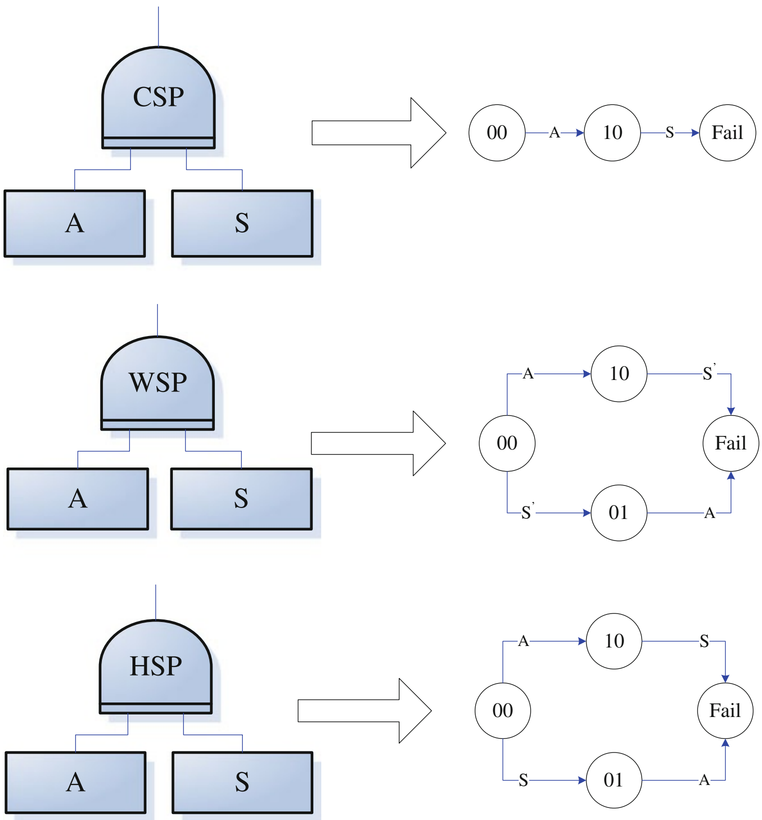


Fig. 26.12 Markov transfer chain of CSP, WSP and HSP

D Failure of P1 or S(S is the substitute for P1);
E Failure of P2 or S(S is the substitute for P2);

So we can get the Markov Transfer Chain as follows:

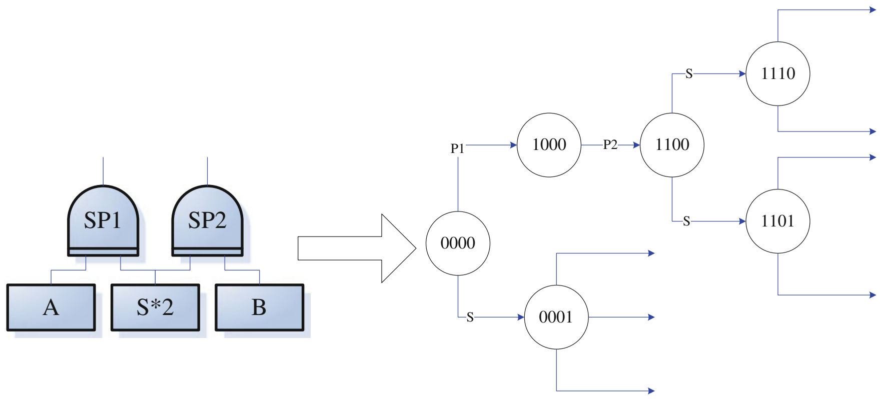


Fig. 26.13 Markov transfer chain of public spare parts pool

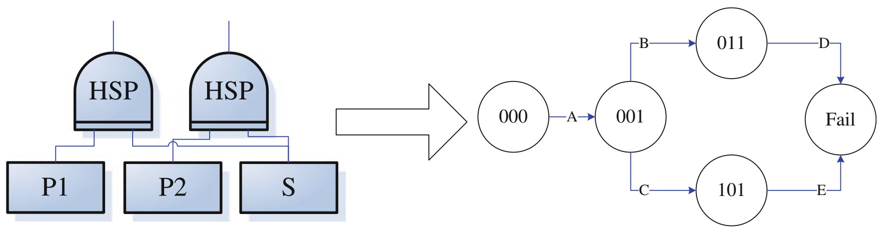


Fig. 26.14 Markov transfer chain of two HSP with one public spare parts pool

26.4 Samples

We assume that oil pump station has three sets of oil pump units: *A*, *B* and *C*. The first two are running normally and the remaining pump unit is the spare part.

26.4.1 Establishment of Dynamic Fault Tree

As shown in Fig. 26.15, there are at least one oil pump unit is in good condition to make sure the normal transport. The oil pump unit is made by pump and electrical machine. The dynamic tree includes the common fault of pump and electrical machine. The events are listed in Table 26.1.

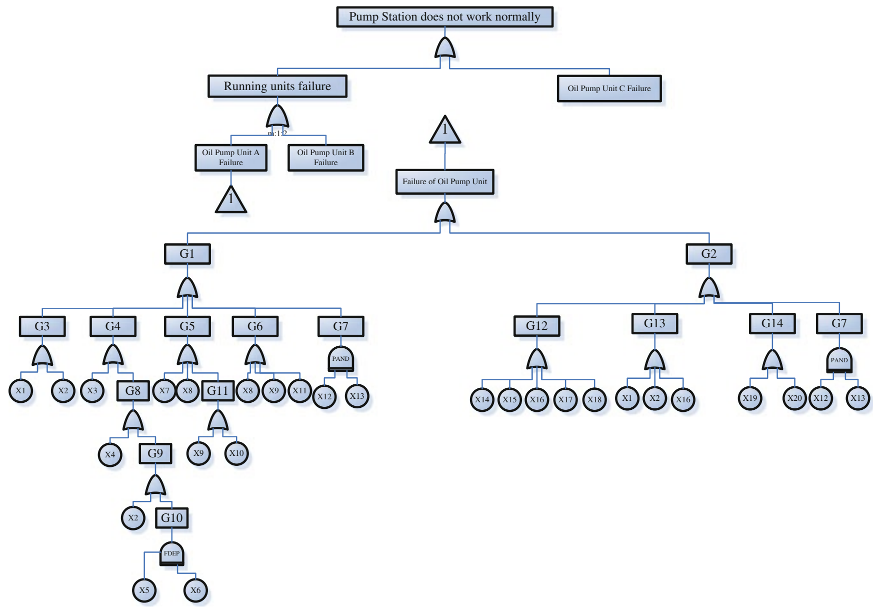


Fig. 26.15 Dynamic fault tree of oil pump units

26.4.2 Static Tree Analysis

By applying the DFLM module search algorithm for dynamic fault tree, we can get the static trees: G3, G5, G6, G11, G12, G13 and G14, and among them G3, G11, G6, G12, G13, G14 are the minimum static sub-tree.

The order of the basic events in Fault Tree is a hotspot in the Fault Tree research. Here I will give a definition

As shown in Figs. 26.16, 26.17, we can get a improved rank criterion [10, 11]:

Criterion 1: “From top to bottom”: the event which in the higher layer of the Fault Tree ranks highly.

Criterion 2: “Repeat priority”: the event which repeats more ranks highly.

Criterion 3: “From left to right”: If the events are the same layer and have the same number of replications, left event has the high rank.

So we can get the retraceable sets of the G3, G5, G6, G11, G12, G13 and G14:

$$G3 : \{\{X2\}, \{\bar{X}2\ X1\}\};$$

$$G11 : \{\{X9\}, \{\bar{X}9\ X10\}\};$$

$$G5 : \{\{X9\}, \{\bar{X}9\ X7\}, \{X8\ \bar{X}7\ \bar{X}9\}, \{X10\ \bar{X}8\ \bar{X}7\ \bar{X}9\}\};$$

$$G12 : \left\{ \begin{array}{l} \{X14\}, \{X15\ \bar{X}14\}\{X16\ \bar{X}15\ \bar{X}14\} \\ \{X17\ \bar{X}16\ \bar{X}15\ \bar{X}14\}\{X18\ \bar{X}17\ \bar{X}16\ \bar{X}15\ \bar{X}14\} \end{array} \right\};$$

Table 26.1 Dynamic fault tree event numbers

Symbol	Event	Symbol	Event	Symbol	Event
G1	Pump failure	G12	Electrical machine overheat	X9	Impurities
G2	Electrical machine failure	G13	Electrical machine axle failure	X10	High flow
G3	Pump axle failure	G14	Shell electrification	X11	Less lubricating oil
G4	Pump leakage	X1	Fatigue or breakage	X12	Exceeding service life
G5	Pump impeller failure	X2	Abrasion	X13	No routine maintenance
G6	Pump bearing failure	X3	Shell breakage	X14	Over high or low voltage
G7	Personal factor	X4	Sealing ring aging	X15	Power imbalance
G8	Sealing failure	X5	High velocity	X16	Overload
G9	Sealing ring failure	X6	Oil overheating	X17	Winding disconnection
G10	Sealing ring groove coking	X7	Cavitation	X18	Poor ventilation
G11	Washout	X8	Corrosion	X19	Grounding lines error
				X20	Insulation aging

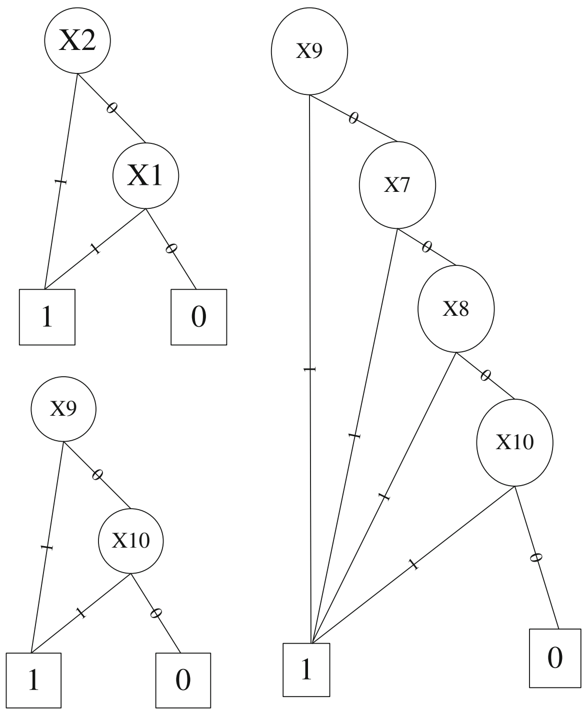


Fig. 26.16 Binary decision diagram of G3, G11 and G5

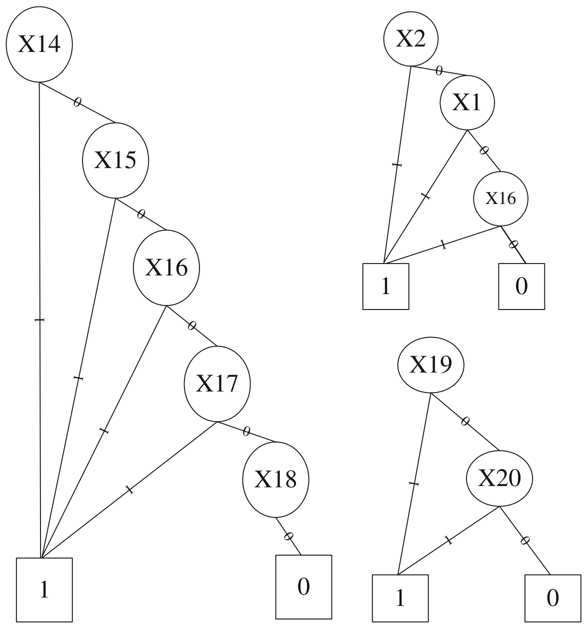


Fig. 26.17 Binary decision diagram of G12, G13 and G14

$$G13 : \{\{X2\}, \{X1\ \bar{X}2\}, \{X16\ \bar{X}2\ \bar{X}1\}\}$$

$$G14 : \{\{X19\}, \{X20\ X\bar{1}9\}\}$$

So we can calculate the fault probability:

$$P(G3) = P(X2) + P(X1)(1 - P(X2)), \tag{26.23}$$

$$P(G11) = P(X9) + P(X10) + (1 - P(X9)), \tag{26.24}$$

$$\begin{aligned} &P(G5) = P(X9) + P(X7)(1 - P(X9)) + P(X8) \\ &(1 - P(X9)) + P(X10)(1 - P(X8))(1 - P(X7))(1 - P(X9)), \end{aligned} \tag{26.25}$$

$$\begin{aligned} &P(G12) = P(X14) + P(X15)(1 - P(X14)) + P(X16) \\ &(1 - P(X15))(1 - P(X14)) + (P(X17))(1 - P(X16)) \\ &(1 - P(X15))(1 - P(X14)) + P(X18)(1 - P(X17)) \\ &(1 - P(X16))(1 - P(X15))(1 - P(X14)), \end{aligned} \tag{26.26}$$

$$\begin{aligned} P(G13) =& P(X2) + P(X1)(1 - P(X2)) \\ &+ P(X16)(1 - P(X1))(1 - P(X2)) \end{aligned} \tag{26.27}$$

$$P(G14) = P(X19) + P(X20)(1 - P(X19)). \tag{26.28}$$

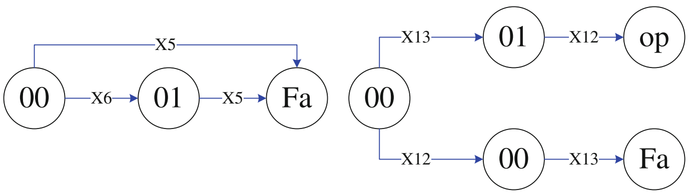


Fig. 26.18 The Markov transfer chain of G7 and G10

26.4.3 Dynamic Tree Analysis

G7 and G10 are the output of the dynamic gate. The Markov Transfer Chain is shown in Fig. 26.18:

So we can get the sequence cut-set of G7 are $\{\bar{X}5\}$ and $\{\overline{X6\,X5}\}$, and G10 is $\{\overline{X12\,X13}\}$. According to the Fokker–Planck equation:

$$P(t) = P(t)Q. \tag{26.29}$$

$P(t)$ indicates the probability of state 0, state 1…state Fa at the time t. $P(t) = [P_0(t), P_1(t), \ldots, P_{Fa}(t)]$. Q indicates the transition probability matrix, and the sum of row in Q equals 0. Take G10 for example, if the failure rate of X12 and X13 are λ_{12} and λ_{13}, the Q is

$$Q = \begin{bmatrix} -(\lambda_{12}+\lambda_{13}) & \lambda_{13} & 0 & \lambda_{12} & 0 \\ 0 & -\lambda_{12} & \lambda_{12} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\lambda_{13} & \lambda_{13} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{26.30}$$

At the initial time $P[0] = [1 \quad 0 \quad 0 \quad 0 \quad 0]$, putting it to the (3–8) then we can the probability of G10.

26.4.4 Comprehensive Solution of Dynamic Fault Tree

Assume that dynamic fault tree $f = f_1(x_1, x_2, \ldots, x_n)(x_1, x_2, \ldots, x_n)$ are the basic events of the dynamic fault tree. The tree includes the dynamic sub-tree $D_1,\ D_2,\ D_i$ and static sub-tree $M_1,\ M_2,\ M_j$. We make $D_1, D_2, \ldots, D_i$ and $M_1, M_2, \ldots, M_j$ as the new bottom events, so the dynamic fault tree can be written as: $f = f_2(x_1, x_2, \ldots, x_1, D_1, D_2, \ldots, D_i, M_1, M_2, \ldots, M_i)$. The probability of new bottom events of $D_1,\ D_2,\ D_i$ and $M_1,\ M_2,\ M_j$ are $P_1,\ P_2,\ P_{i+j}$. Then we can solve the fault probability of top event.

26.5 Conclusions and Future Work

This paper introduces the basic theory of dynamic fault tree, followed by an example in which this method is used to analyze a pump unit. The analytical method includes Binary Decision Diagram method and Markov Transfer Chain method. BBD method is mainly basing the ite structure, while the order of the basic events is difficult and of great focus. The paper gives three criterions in how to rank basic events. The example given shows that the sequencing problem becomes easy and effective under the new criterions. In allusion to the whole dynamic tree, paper puts forward that the transforming between basic event and output events of gate.

Although this paper gives a methodto address ranking basic events the accuracy needs further improvement. The expansion of Markov Transfer Chain is still a critical issue, some researchers are trying to solve the problem of how to deal with the long Markov Transfer Chain. So the two difficult points are still worthy areas for study.

Acknowledgments This project is supported by National Natural Science Foundation of China (Grant No. 51005247); National High Technology Research and Development Program of China (Grant No. 2008AA06Z209); Beijing Nova Program(Grant No.2010B068); Special Items Fund of Beijing Municipal Commission of Education; Research Funds Provided to New Recruitments of China University of Petroleum-Beijing (Grant No. QD-2010-01)and China University of Petroleum-Beijing.

References

1. Dugan JB, Bavuso SJ, Boyd MA (1992) Dynamic fault-tree models for fault-tolerant computer systems. IEEE Trans Reliab 41(3):363–377
2. Rauzy AB (2011) Sequence algebra, sequence decision diagrams and dynamic fault trees. Reliab Eng Syst Saf 96(7):785–792
3. Rao KD, Gopika V, Sanyasi Rao VVS, Kushwaha HS, Verma AK, Srividya A (2009) Dynamic fault tree analysis using Monte Carlo simulation in probabilistic safety assessment. Reliab Eng Syst Saf 94(4):872–883
4. Huang CY, Chang YR (2007) An improved decomposition scheme for assessing the reliability of embedded systems by using dynamic fault trees. Reliab Eng Syst Saf 92(10):1403–1412
5. Zhang HL, Fu J, Zhang HY, Wang YY (2011) Dynamic fault tree analysis method based on isomorphic node. Comput Eng Des 32(1):1–4
6. Sun Y (2007) DFT's application in reliability analysis of automatic train supervision system in subway. [MS Thesis]. Beijng: Beijing Jiaotong University, 2007
7. Manian R, Coppit DW, Sullivan KJ, Dugan JB (1999) Bridging the gap between systems and dynamic fault tree models. In: Proceedings annual reliability and maintainability symposium, p 105–111
8. Ren Y, Dugan JB (1998) Design of reliable system using static and dynamic fault trees. IEEE Trans Reliab 47(3):234–244
9. Bryant RE (1986) Graph-based algorithms for boolean function manipulation. IEEE Trans Comput 35(8):677–691

10. Tao YJ, Dong DC, Ren peng (2009) Fault tree analysis based on binary decision diagrams. Railway Comput Appl 18(9):4–7
11. Sun Y, Du SG (2008) A novel ordering method of binary decision diagram. In: Proceedings of the international conference on management science and engineering, p. 299–304
12. Gulati R, Dugan JB (1997) A modular approach for analyzing static and dynamic fault trees. In: Proceedings annual reliability and maintainability symposium, p 57–63

Chapter 27
Operation Reliability Assessment Based on Running Condition Information for Large Machinery

Z. He and G. Cai

Abstract Traditional reliability assessment methods predominantly based on probability and statistical theories need a large sample size to get the general, overall estimates for the identical units. However, the sample is often insufficient for large machinery, traditional reliability assessment methods became ineffective. More importantly, traditional reliability assessment cannot reflect the individuality of the running equipment. To overcome these deficiencies, a new approach based on running condition information to assess the operation reliability of large machinery is proposed. First, according to "short board effect", the critical and weak component of the equipment is determined. Second, based on quantitative damage diagnosis, a mapping function called reliability membership function is built between equipment damage severity and a new reliability index which is defined as membership reliability. Third, quantitative damage diagnosis is employed to achieve the damage severity feature of the key component. Finally, by substituting the feature into the function, the value and change trend of the membership reliability is calculated and the equipment's operation reliability assessment is gained. The application to the reliability assessment of the bearing for large machinery demonstrated the proposed method is reasonable and effective. Moreover, the proposed approach provides a new way to the operation reliability assessment for large machinery.

Z. He (✉) · G. Cai
State Key Laboratory for Manufacturing System Engineering,
Xi'an Jiaotong University, Xi'an 710049, China
e-mail: hzj@mail.xjtu.edu.cn

G. Cai
e-mail: caigaigai.2008@stu.xjtu.edu.cn

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_27,

27.1 Introduction

Traditional reliability assessment approaches, primarily based on probability and statistical theories, need a large sample of historical time-to-failure data to get the general and overall estimates for the entire population of identical units [1]. However, since the sample size is often small for large machinery, traditional reliability assessment approaches have become difficult to be used. Even more importantly, for the equipment that is running, the main focus is its operation reliability, which is different from its designed reliability. Accordingly, how to assess the reliability of large machinery that is running, i.e., operation reliability, in small sample size becomes an urgent problem needs to be studied.

The dynamic signals from machinery will present impulse response when damage appears. Take rolling bearing for example, when damage occurs in the bearing, its dynamic response signal will present oscillation attenuation by the impact pulse excitation at the damaged place. Thus, machinery running condition information can effectively reflect the characteristic of dynamic operation and machining accuracy of the unit, and then reveal the performance degradation of the equipment. Moreover, useful information for reliability assessment can be extracted from the running condition information. In recent years, there has been a great amount of research being conducted on reliability modeling and assessment by using running condition information [2–6]. Most of this reasearch focuses on taking advantage of reliability assessment models such as proportional hazard model and logistic model, and so on to build the relationship between equipment reliability and running condition information. These methods, however, generally require a large sample of time-to-failure data and running condition information to achieve reliability assessment. Unfortunately, samples of the required size and historical time-to-failure data for large machinery are often rare. These methods therefore are extremely difficult to apply.

In this paper, a new reliability assessment method based on running condition information for large machinery using a small sample size is proposed. First, the critical and weak component of the equipment is determined according to "short board effect". Second, a new reliability index called membership reliability that can describe the membership degree of the running state of the equipment belongs to the state of running safely and reliably is introduced. By taking advantage of the result of the quantitative damage diagnosis, a mapping function, named the reliability membership function, between damage severity and membership reliability is established. Third, the damage severity feature of the key component is gained with the aid of quantitative damage diagnosis. Finally, by substituting the damage severity feature into the function, the value and the change trend of the membership reliability can be calculated, the operation reliability assessment for the equipment using a small sample size is gained. The proposed method is applied to assess the operation reliability of the key component of large machinery. This application indicates the proposed method is reasonable and effective.

The presented study provides a new idea and method for operation reliability assessment of large machinery in small sample size.

The rest of the paper is organized as follows. The reliability assessment method based on the running condition information is described in Sect. 27.2. In Sect. 27.3, the proposed method is used for rolling bearing operation reliability assessment of large machinery. Conclusions are given in Sect. 27.4.

27.2 The Proposed Method

Given the "short board effect" the reliability of equipment such as roller bearing, the key is the shortest piece on board, rather than the longest piece. Rolling bearing plays a critical role in large machinery, but also is one of the most easily damaged components [7]. According to statistics, up to 30 % of the failure of the rotating machines is caused by the rolling bearings. The operation reliability of the rolling bearing directly influences the operation safety of the machinery. Hence, we choose bearings in large machinery as the critical and weak components to implement operation reliability assessment. By evaluating the operation reliability of rolling bearings, the operation reliability assessment of large machinery can be made.

27.2.1 Quantitative Damage Diagnosis for Rolling Bearings Based on Running Condition Information

The shock pulse method (SPM) was developed by SPM Instrument AB in the early 1970s in Sweden. It relies on the fact that when the ball or roller makes contact with a damaged area of the raceway or with debris in the rolling bearing, there will be impulse of vibration [8]. Using the SPM, the maximum normalized shock value (dB) gives an indication of bearing condition. It is defined as [9],

$$\mathrm{dB} = 20\log\frac{2000 \times SV}{N \times D^{0.6}}, \tag{27.1}$$

in Eq. (27.1) N denotes the rotating speed of the bearing, D denotes the inner diameter, and SV is the shock value, which should be calculated by demodulation of bearing vibration signals, it is the vibration velocity value of the signal, and its unit is m/s. The bearing running state can be estimated with the dB value [9]

$$\begin{cases} 0 \le \mathrm{dB} \le 21 & \text{Healthy bearing;} \\ 21 < \mathrm{dB} \le 35 & \text{Weak damaged;} \\ 35 < \mathrm{dB} \le 60 & \text{Heavy damaged.} \end{cases} \tag{27.2}$$

Within the established boundaries of three condition zones, the SPM has been widely used as a quantitative method for detecting the operating condition and damage severity of rolling bearing [7, 11–16]. The SPM combined with demodulation method is a useful method of estimating bearing running condition. The maximum normalized shock value gives an indication of bearing damage condition. However, the response signal is often weak and useful information is always submerged by strong environmental noise, direct demodulation may estimate the shock value in the SPM ineffectively. Improved redundant second generation wavelet transform (IRSGWT) is applied to preprocess the vibration signal of rolling bearing before demodulation. Its principle can refer to Ref. [17].

The procedure of the quantitative damage diagnosis method based on IRSGWT and the SPM is as follows:

1. Apply redundant second generation wavelet transform to decompose the vibration signal of the rolling bearing, then calculate the normalization factors of different frequency band;
2. Using the normalization factors to improve the decomposition result of the redundant second generation wavelet transform;
3. Apply Hilbert spectrum to the decomposition result of the IRSGWT, get the amplitude envelope of different frequency bands;
4. Compute the fault characteristic frequencies of the bearing;
5. Calculate the standard dB value of the fault characteristic frequencies in the frequency band which contains the predominant impulse;
6. Choose the maximum values to identify the damage severity of the bearing.

27.2.2 Operation Reliability Assessment Based on Quantitative Damage Diagnosis

In order to realize the operation reliability assessment of large machinery on basis of quantitative damage diagnosis, we focus on building the mapping function between the reliability index and damage severity. In fuzzy math, a fuzzy set is determined by its membership function, and the membership function is an explicit function [10]. In this paper, we introduce a new reliability assessment index called membership reliability to assess the operation reliability of machinery using running condition information. By establishing the reliability membership function between membership reliability and damage severity feature, the reliability assessment of the equipment will finally be achieved.

First, a reliability set A is defined as “Equipment runs safely and reliably”. Second, the membership reliability that could quantitatively describe the membership degree of the running state of the equipment belonging to set A is proposed. Here, we denote the membership reliability as R_A. Its value range is from 0 to 1, and it can map the physical characteristic parameters that represent the different running states of the equipment to a dimensionless range from 0 to 1. Third, the mapping function, i.e. the reliability membership function, $R_A(M)$, between damage severity and membership reliability is established. The damage severity feature is denoted as M. The greater value M is, the worse the damage is and lower the reliability of the equipment is. That is, the membership degree of the equipment runs safely and reliably is smaller. Hence, in the present study, we introduce Eq. (27.3) to express its reliability membership function $RA(M)$ by studying the performance degradation of machinery. $R_A(M)$ is as follows:

$$R_A(M) = \begin{cases} 1 & 0 \le M \le a_1 \\ e^{-k(M-a_1)^2} & a_1 < M \le a_2 \\ 0 & M > a_2 \end{cases}. \tag{27.3}$$

The value range of $R_A(M)$ is $0 \le R_A(M) \le 1$. In Eq. (27.3), the values of constants k and a_1, a_2 can be calculated from quantitative damage diagnosis.

Take rolling bearings into consideration, according to Eq. (27.2), we define the corresponding value of the membership reliability to be 1 if $0 \le \text{dB} \le 21\text{dB}$. And when we get $21 < \text{dB} \le 60\,\text{dB}$, the bearing's performance is described as degraded, and its reliability declined. The bearing's running state is between slight damage and serious damage when $\text{dB} = 35$. Consequently, based on the "either this or that" mathematical principle, we set the membership reliability equal to 0.5 in case $\text{dB} = 35$. Meanwhile, 0.5 is set to be the threshold to judge whether the equipment is running reliably, i.e., if the calculated membership reliability is below 0.5, the equipment is considered as unreliable. When measured in decibels, the shock value of bearing failure life is 60 dB. Thus if the shock value for a bearing is greater than 60dB, its membership reliability is set to be 0. Then, k, a_1 and a_2 are calculated. Finally, the reliability membership function for rolling bearing is as follows,

$$R_A(M) = \begin{cases} 1 & 0 \le M \le 21\,\text{dB} \\ e^{-0.0035(m-21)^2} & 21 < M \le 60\,\text{dB} \\ 0 & M > 60\,\text{dB} \end{cases} \tag{27.4}$$

Here, $R_A(M) = 0.5$ is chosen as the reliability threshold for rolling bearing depending on engineering experience. If $R_A(M) = 1$, it indicates that the equipment running state belongs to A, and the equipment is in good working condition. While the calculated result for $R_A(M)$ is a value between 0.5 and 1, it indicates that the reliability of the equipment has declined compared with the good condition, the smaller the value is, the lower the reliability. And if the value range for $R_A(M)$ is $0 < R_A(M) < 0.5$, that the equipment is considered to be extremely damaged, and its reliability falls sharply. To avoid a serious accident, it must be overhauled immediately. And if the calculation result for $R_A(M)$ is 0, it indicates that the equipment is totally damaged and completely unreliable.

27.3 Engineering Applications

In this section, the proposed method is applied to assess the operation reliability of electric locomotives and a dust removal fan used in steelmaking.

27.3.1 Bearing Reliability Assessment of Electric Locomotive

As a key transmission component of electric locomotives, wheel bearings operate under severe environment for a long time.

According to "short board effect", rolling bearing is considered as the weak board of electric locomotives. The proposed method is used to assess the operation reliability of rolling bearings of electric locomotives under different running states. The bearing type is 552732QT and its geometric parameters are listed in Table 27.1. The vibration signals are picked up at a constant shaft speed of 390 r/min and digitized at a sampling frequency of 12.8 kHz. The characteristic frequency of the outer-race defects is calculated to be at 46.88 Hz based on the geometric parameters and rotating speed of the bearing.

Figure 27.1 shows the bearing vibration signals under different running states, state 1 (Fig. 27.1a), state 2 (Fig. 27.1b), and state 3 (Fig. 27.1c).

First, to extract the useful information from the vibration signals accurately, IRSGWT is used to preprocess the vibration signals under different running states. Figures 27.2, 27.3, and 27.4 show the selected frequency bands $r^{(1)}$ that contain the predominant impulse after 4-level decomposition of the vibration signals using IRSGWT under different running states. Their Hilbert spectrums are shown in Figs. 27.5, 27.6, and 27.7, respectively.

Based on Figs. 27.5, 27.6, 27.7 above, the maximum normalized shock values (dB) of the bearings are calculated according to Eq. (27.1). For state 1 it is 16.5 dB. And it is 31.6 dB for state 2. For state 3 it is 49.8 dB.

Substitute the results of quantitative damage diagnosis for rolling bearings into Eq. (27.4), the membership reliability for $M = 16.5$ is $R_A(16.5) = 1$. For $M = 16.5$, it is $R_A(31.6) = 0.6749$, and it is $R_A(49.8) = 0.549$ for $M = 49.8$. The operation reliability assessment results for different states are listed in Table 27.2.

Table 27.1 The geometric parameters of the tested bearing

Roller diameter	34 mm
Inner race diameter	160 mm
Outer race diameter	290 mm
Number of rolling elements z	17
Contact angle α	0°

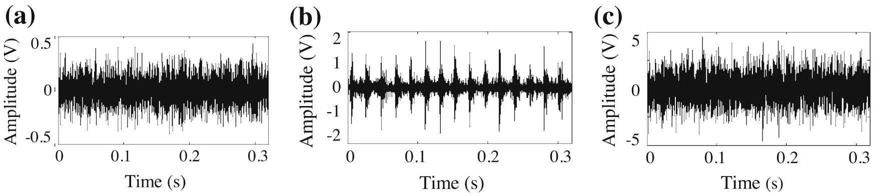


Fig. 27.1 The vibration signals of bearing under different running states: state 1 (**a**), state 2 (**b**), and state 3 (**c**)

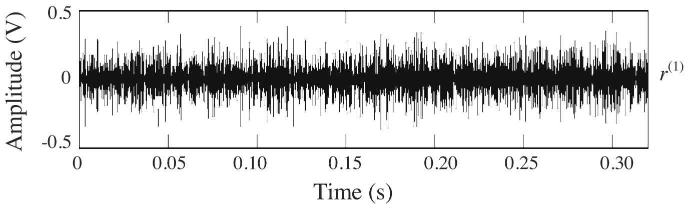


Fig. 27.2 $r^{(1)}$ of the decomposition result using IRSGWT of the vibration signal for state 1

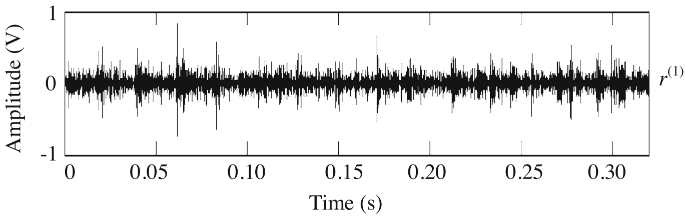


Fig. 27.3 $r^{(1)}$ of the decomposition result using IRSGWT of the vibration signal for state 2

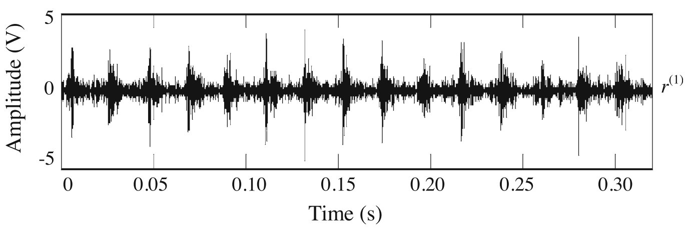


Fig. 27.4 $r^{(1)}$ of the decomposition result using IRSGWT of the vibration signal for state 3

Afterwards, the wheel bearings are removed for analysis, the wheel bearing is in good condition in state 1. There is a slight rub on the outer race of the bearing in state 2. In state 3, there is severe spalling on the outer race of the testing bearing (Fig 27.8).

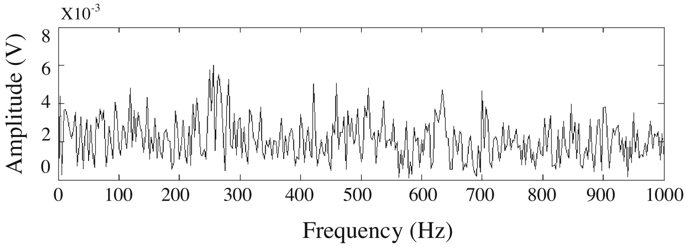


Fig. 27.5 The Hilbert spectrum of $r^{(1)}$ for state 1

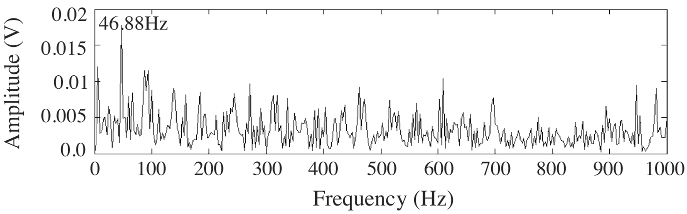


Fig. 27.6 The Hilbert spectrum of $r^{(1)}$ for state 2

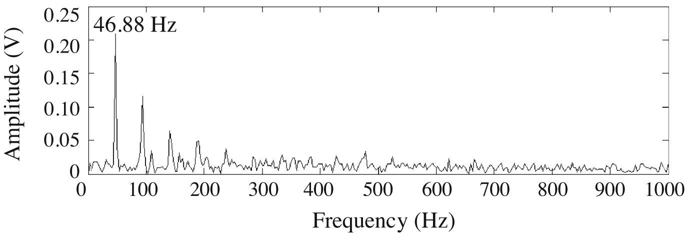


Fig. 27.7 The Hilbert spectrum of $r^{(1)}$ for state 3

Table 27.2 Operation reliability assessment for bearings using the proposed method

	State 1	State 2	State 3
$M(dB)$	16.5	31.6	49.8
$R_A(M)$	1	0.6749	0.0549
Reliability assessment result	Equipment runs safely & reliably	Equipment reliability declined, and the equipment should be overhauled in a short time	Equipment is running unreliably, and must be overhauled immediately

Fig. 27.8 Actual photos for the locomotive rolling bearings in different states

27.3.2 Bearing Reliability Assessment of Steelmaking Dust Removal Fan

As a key equipment of steelmaking furnace, steelmaking dust removal fan runs in severe environments such as, high temperature, and its transmission component will can be easily be damaged, and this will seriously affect the production.

In order to assess the operation reliability of the steelmaking dust removal fun in a steel works, we choose rolling bearing in the load side of the steelmaking dust removal fan as its key component to assess its operation reliability. The vibration signals were measured by velocity sensors mounted on the bearing seat and digitized at a sampling of 5,120 Hz. The bearing designation is 22226EMW33C3. The rotation frequency of the fan is 12.5 Hz. The fault characteristic frequencies of the bearing are calculated and shown in Table 27.3.

Figure 27.9 shows the vibration signal of the bearing. Its frequency domain signal is shown in Fig. 27.10. There is no evident fault characteristic frequency in Fig. 27.10.

A 3-level decomposition of the vibration signal is performed by using IRS-GWT. Figure 27.11 represents the Hilbert spectrum of the third frequency band which has the predominate impulse after 3-level decomposition. From Fig. 27.11, the maximum normalized shock value (dB) of the bearing is calculated to be 28.9328 dB. Finally, the dB value is substituted into Eq. (27.4), the membership reliability for the bearing is $R_A(M) = R_A(28.9328) = e^{-0.0035(28.9328-21)} = 0.8023$.

Table 27.3 The characteristic frequencies for the load side bearing of the fan

Bearing fault characteristic frequencies	Frequency (Hz)
The characteristic frequency of the retainer	5.4375
The characteristic frequency of the outer-race	39.9125
The characteristic frequency of the roller	95.5375
The characteristic frequency of the inner-race	129.4625

Fig. 27.9 The vibration signal of the load side bearing of the fan

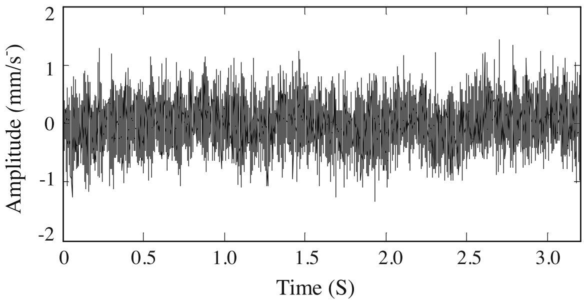


Fig. 27.10 The Fourier transform of the signal of Fig. 27.9

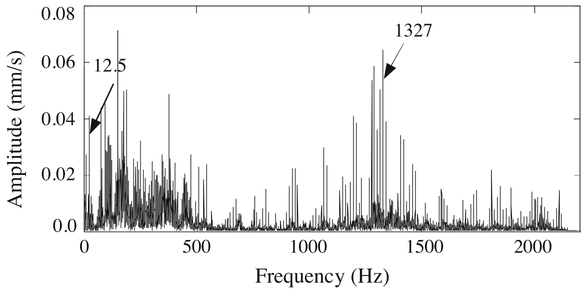
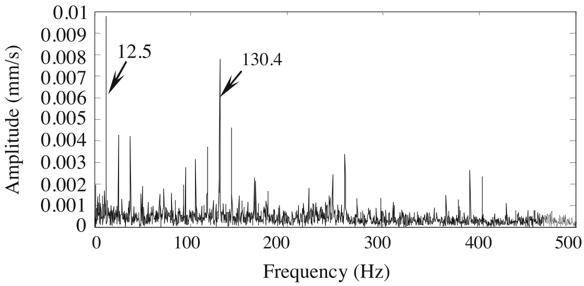


Fig. 27.11 The third frequency band of the decomposition result using IRSGWT



The calculated result indicates the dust removal fan's reliability has is declined compared with the good condition, and it should be overhauled in a short time. Later, the dust removal fan was overhauled and a slight scratch is found in the inner-race of the bearing, as shown in Fig. 27.12.

Engineering application results demonstrate that the proposed method is reasonable and effective for operation reliability assessment of rolling bearings. Via the value and change trend of the membership reliability, the operation reliability for large machinery is evaluated. Furthermore, the proposed method also provides a promising way for operation reliability assessment of large machinery.

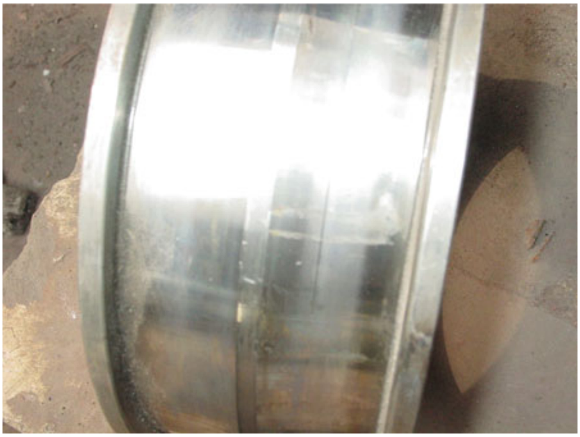


Fig. 27.12 A slight scratch on the inner-race of the bearing

27.4 Conclusion

This paper presents a new method for operation reliability assessment for large machinery based on running condition information. In the proposed method, IRS-GWT together with Hilbert transform and the SPM can realize quantitative damage diagnosis of rolling bearing accurately. By the newly proposed reliability index, membership reliability, the membership degree of the equipment belonging to the condition of operating safely and reliably can be described. By establishing the reliability membership function between membership reliability and damage severity, the membership reliability is calculated. According to the value and trend of the membership reliability, operation reliability assessment for the machinery is realized. In the proposed method, the physical characteristic parameters that represent the different running states of the equipment is mapped to a non-dimensional range from 0 to 1, thus the running state of the equipment in different running conditions can be compared with each other. The application result demonstrated that the proposed method is reasonable and effective for operation reliability assessment of large machinery. The proposed method in this paper provides a promising way for large machinery operation reliability assessment in case of small sample and non-failure data.

Acknowledgments This work was supported by the key project of National Nature Science Foundation of China (No.51035007), the National Basic Research Program of China (Grant no.2011CB706805).

References

1. Heng ASY (2009) Intelligent prognostics of machinery health utilizing suspended condition monitoring data. Queensland University of Technology, Queensland
2. Lin C, Tseng H (2005) A neural network application for reliability modeling and condition-based predictive maintenance. Int J Adv Manuf Technol 25(1):174–179

3. Liao H, Zhao W, Guo H (2006) Predicting remaining useful life of an individual unit using proportional hazards model and logistic regression model. In: IEEE proceedings annual reliability and maintainability symposium, pp 127–132
4. You M-Y, Lin L, Guang M (2010) Two-zone proportional hazard model for equipment remaining useful life prediction. J Manuf Sci Eng 132:041008-1–041008-6
5. Feng D, Zhengjia H, Yanyang Z et al (2009) Reliability assessment based on equipment condition vibration feature using proportional hazards model. Chin J Mech Eng 45(12):89–94
6. Chen B, Chen X, Li B (2011) Reliability estimation for cutting tools based on logistic regression model using vibration signals. Mech Syst Signal Process. doi:10.1016/j.ymssp.2011.03.001
7. Li Z, He Z, Zi Y et al (2008) Bearing condition monitoring based on shock pulse method and improved redundant lifting scheme. Math Comput Simul 79:318–338
8. He Z, Zi Y, Zhang X (2007) Modern signal processing and engineering application. Xi'an Jiaotong University Press, Xi'an
9. The shock pulse method for determining condition of anti-friction bearings. SPM Technical Information, Sweden, *SPM Instruments AB* http://www.conditionenergy.com/SPM_vs_Vibration.pdf
10. Zimmermann HJ (1985) Fuzzy set theory and its application. Kluwer-Nijhoff Publishing, Hingham, pp 11–37
11. Butler DE (1984) The shock pulse method for the detection of damaged rolling bearing. NDT Int 6(1973):92–95
12. Mathew J, Alfredson RJ (1984) The condition monitoring of rolling element bearing using vibration signal. Trans ASME J Vibr Acoust Stress Reliab Des 106:145–154
13. Morando LE (1988) Measuring shock pulses is ideal for bearing condition monitoring. Pulp Paper 62:96–98
14. Prabhu R (1996) Rolling element diagnostics. In: Proceeding of the Indo-US symposium on emerging trends in vibration and noise engineering, New Delhi, pp 311–320
15. Morando LE (1998) Bearing condition monitoring by the shock pulse method. Iron and Steel Engineers:40–43 http://digital.library.aist.org/pages/PR-PM1298-5.htm
16. Tandon N, Yadava GS, Ramakrishna KM (2007) A comparison of some condition monitoring techniques for the detection of defect in induction motor ball bearings. Mech Syst Signal Process 21:244–256
17. Li Z (2008) Research on second generation wavelet theory and its applications in fault diagnosis. Xi'an Jiaotong University Press, Xi'an

Chapter 28
Proactive Fleet Health Monitoring and Management

M. Monnin, J. B. Leger and D. Morel

Abstract Achieving fleet-wide management and decision-making support raises challenges that still have to be addressed. Toward this end, new methodologies, methods and tools are required to identify the related health indicators, support continuous monitoring, enable predictive diagnosis, and provide suitable prognosis facilities in order to carry out proactive health management. In this paper, an approach to proactive fleet health monitoring and management is proposed. The underlying modeling methods are introduced as well as health monitoring facilities for proactively reacting to system performance drift.

28.1 Introduction

For supporting maintenance personnel in their decision making process, only condition monitoring is not sufficient; diagnosis and prognostics must be considered in order to reach proactive strategies [1]. Such predictive diagnosis and prognostics is motivated by the need for manufacturers and other operators of complex systems to optimize equipment performance and reduce costs and unscheduled downtime. As a result, system/plant health monitoring system is developed driving maintenance activities at the system level based on the complex equipment health status. In addition, large engineering systems such as power

M. Monnin (✉) · J. B. Leger · D. Morel
PREDICT, Technopole de Brabois19 Avenue de la Foret de Haye—CS 10508,
54519 Vandoeuvre Les Nancy, France
e-mail: maxime.monnin@predict.fr

J. B. Leger
e-mail: jean-baptiste.leger@predict.fr

D. Morel
e-mail: david.morel@predict.fr

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_28,

plants, ships, and aircraft have multiple systems, subsystems, and equipment of a mechanical, electrical, electronic, and software nature which follow different rates and modes of failure [2] and for which behavior can vary across different phases of their lifecycle [3]. When operated as a fleet, mission readiness and maintenance management issues are raised. Thus supporting proactive maintenance at the fleet level should answer questions such as: given a mission of duration X days, what percentage of units assigned to that mission are able to complete the mission without a critical failure? Or does this equipment condition state have already been reach by another equipment of the fleet?

The need for decision making support within complex, dynamic, and large-scale environments is real. Existing approaches tend to fill this need by providing advanced condition monitoring-based approaches. Such approaches start from equipment condition monitoring to reach the fleet. A proactive fleet management system is not really operational on site because the three key processes (prognosis, diagnosis, and monitoring) are limited with regard to the large scale of failures and degraded states at the scale of the fleet. Thus, proactive fleet management raises issues such as how to process large amount of heterogeneous data (i.e., interpretable health indicator assessment), how to facilitate diagnosis of a heterogeneous fleet of equipment (i.e., several technologies and usage), and how to provide key users with an efficient decision making support throughout the whole asset lifecycle. Toward this end, a complete guide (model, method, and tool) is needed to support the modeling and the understanding of such systems at the scale of the fleet [4]. In this paper, a unified and integrated approach is proposed. This approach is described as unified because it relies on a global methodological approach for which each step is supported by means of modeling methods and tools. Integrated because it is supported by means of PREDICT's software platform that delivers to the users a consistent framework by including a multi-technologies/methods toolbox that support a broad range of algorithm for crossing from one level to another (i.e., reconciliation, consolidation, and aggregation). PHM (1) Engineering Maintainability Study System Engineering Optimisation Expertise Feedback Commissioning Fleet Mgt Engineering Design Operation Execution KBS (2) Prognosis Decision Support Monitoring & Diagnosis (1) PHM: Prognostic & Health Management (2) KBS: Knowledge-Based System

28.2 Unified and Integrated Approach for Effective Proactive Management

28.2.1 Closed-Loop for Engineering Improvement

The proposed approach is built around a global methodology presented in Fig. 28.1. The methodological approach allows guiding the proactive strategy definition all along the asset lifecycle. On the one hand, starting from the early

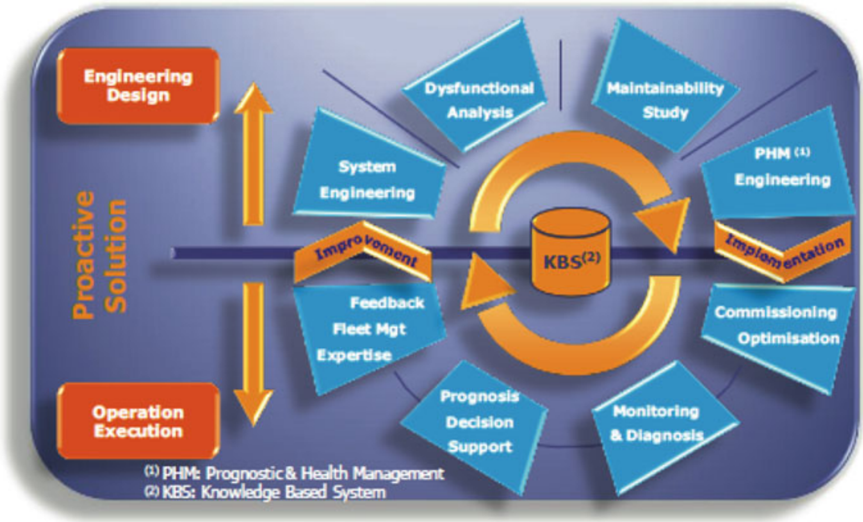


Fig. 28.1 A global methodological approach

engineering phase, it allows a better understanding of equipment and system behavior by means of integrating the model of the global malfunctioning description of the system (deviation, degradation behavior, and causality) with its functioning representation. Functional analysis, FMECA, and HAZOP analysis methods are supported in order to formalize the corresponding knowledge [5]. This modeling approach allows targeting the needs of monitoring by knowing predominant failure modes for instance. Moreover, the definition of the system functioning in relation to its environment helps identifying the best suited monitoring techniques to be applied according to the features to be monitored.

28.2.2 Fleet-Wide Approach for a More Global Feedback

On the other hand, this approach enables fast proactive health monitoring and assessment by a multi-technologies/methods toolbox for data treatment. That enables early drift in monitored features to be detected and analyzed. Then the corresponding knowledge (i.e., functioning and malfunctioning models) is incrementally built in the knowledge-based system. In both cases, the approach provides a consistent framework that enables coupling data and models to support diagnosis, prognostics, and expertise through a global and structured view of the system and enhances understanding of abnormal situations (i.e., system health monitoring level—Fig. 28.2).

In addition to individual (i.e., system level) modeling and health monitoring, it is necessary to provide fleet-wide facilities. Toward this end, the platform

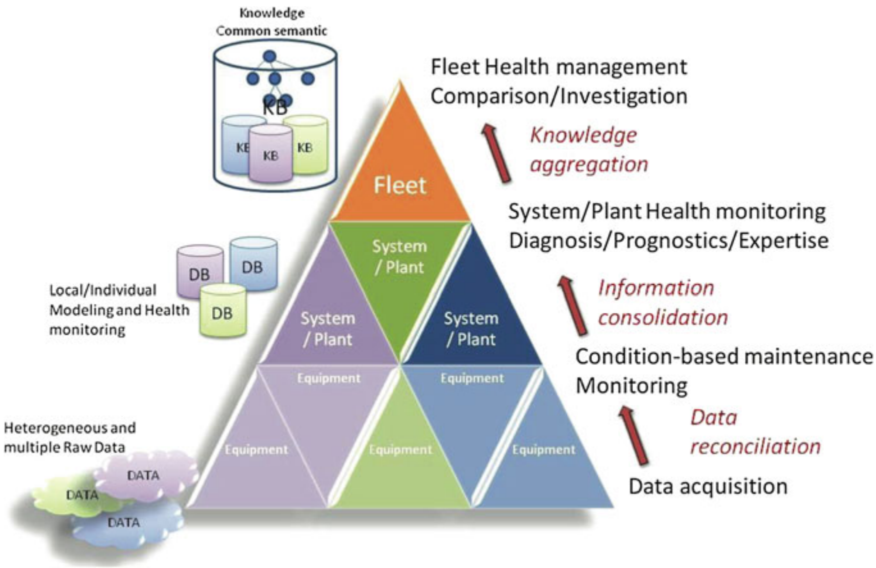


Fig. 28.2 Hierarchical approach of the proactive fleet management

integrates semantic modeling that allows gathering and sharing at fleet wide, data, models, and expertise (e.g., diagnostic results) from the different systems and equipment.

By considering domain specific attributes, data and information can be analyzed through comparison according to different points of views (e.g., compare condition index of similar equipment in the different location, compare different system health trend for the same mission…). Moreover, the fleet facilities integrate aggregation toolbox for providing managers and engineers with a complete summary of information in order to keep them updated regarding both the global health of the fleet and the present status of their maintenance efforts (i.e., fleet health management level—Fig. 28.2).

28.2.3 Closed-Loop and Fleet-Wide for PHM Performances

Moreover, the platform integrates a specific workflow that provides the user with advanced functions at each step of a solving problem process. It allows tracking each decision and analysis from drift detection to proactive action (i.e., before the failure occurs). It allows, alarms status (occurred, analyzed, diagnosed…) to be known along with the visualization of associated degradations or deviation (according to the corresponding modeling). The treatment of the alarms and events is managed through a workflow enabling the traceability of work evolution (i.e., comment, diagnosis, action, documentation). In addition to that, according to the different steps of the workflow, advanced support is provided for alarms and events analysis. One can have access to causes and consequences linked to event from the knowledge database and run through the causal relations.

Thus, each step of the process could benefit from the fleet-wide return of experience. The structured knowledge-based system allows similar situations to be retrieved and compared for diagnosis purpose and prognostics [6]. By this way, the knowledge base is continuously improved and refined allowing to continuously enhancing efficiency and relevance of the fleet-wide health assessment and management. The corresponding return of experience is automatically built and structured.

28.3 Industrial Application

The industrial application demonstrates how the preceding concepts are embedded in a commercial application [4, 5] developed by PREDICT. The example presents abnormal situation analysis helping using similar case retrieval within the fleet. The aim of the analysis is to anticipate failure, i.e., to perform predictive diagnosis.

28.3.1 Case Description

Diesel engines are critical onboard component of ship. In many cases they provide both propulsion of the ship and electrical power within many possible configurations. Avoiding blackout is of primary concerns and marine diesel engine monitoring and maintenance tend to benefit from advanced technology. Indeed, because embedded maintenance facilities are limited, a better knowledge of the engine health condition will allow to better drive maintenance actions needed when ships are in port.

From a predictive diagnosis point of view, existing alarm monitoring systems are not sufficient since they do not allow failure to be anticipated. Once the alarm occurs, the remaining time to failure is too short for preventing it. Moreover, the cause identification of such alarms must be analyzed subsequently. For the sake of illustration, let us consider cylinder temperatures for diesel engines. In normal conditions, the cylinders temperatures are changing in a similar way. Thus, a health indicator of abnormal behavior shall be built by detecting any evolution of one of the temperatures disconnected from the rest of the set of temperatures. Figure 28.3 illustrates temperatures measurement evolution of a diesel engine. Two behaviors are highlighted on the graph. The first behavior, labeled A, shows a normal situation where the temperatures are correlated despite one of them is a couple of degrees below. The second behavior, labeled B, shows a decorrelation of the lowest signal.

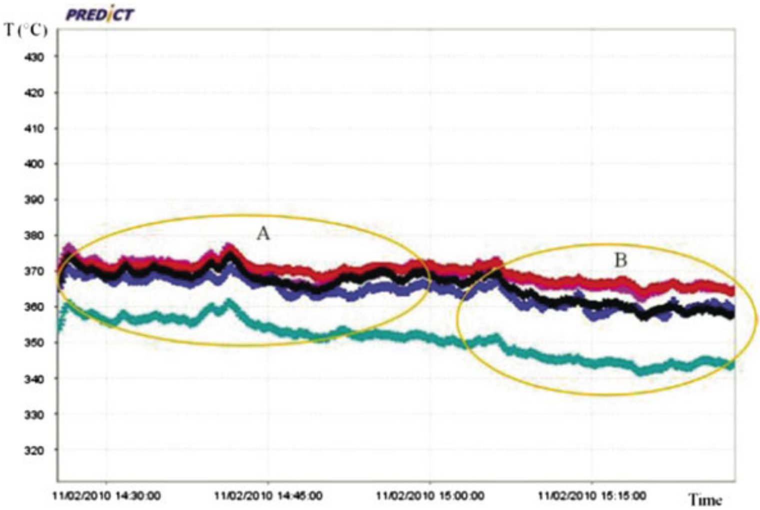


Fig. 28.3 Zoom over a one-hour period of cylinder temperature measurement

Such data trend analysis, even if coupled with a detection process, will not anticipate failure as the, contextual information for the abnormal behavior that enable the understanding (i.e., diagnostic) of the behavior are missing. For the purpose of diagnosis, the platform allows characterizing an event by building a snapshot of relevant variables within a defined time period before and after the event. Obtaining such, contextual information for of an abnormal behavior helps experts during the interpretation of the system behavior during the event.

The platform allows similar past situations—i.e., context of an event with the associated diagnosis—to be retrieved in order to facilitate diagnosis by comparison. For a fleet of engines, the underlying structured knowledge, allows gathering situations from the ship databases (cf. Fig. 28.2). Thus, regarding the current situation to be investigated, the expertise process benefits from fleet-wide return of experience. The knowledge model provides different criteria in order to gather past situation information.

For instance on the basis of the abnormal behavior detected through the cylinder temperature monitoring, one can ask for similar situation (Fig. 28.4) regarding for instance the engine configuration (V or in-line), the number of cylinder 6/8/12 or even their usage (e.g., propulsion, electric power supply).

Thus, fleet-wide knowledge can be shared in order to enhance proactive management. Past experiences can be retrieved and compared in a timely manner. Coupled with monitoring and early detection, the expertise and diagnosis processes are improved to carry out drift behavior interpretation and failure anticipation.

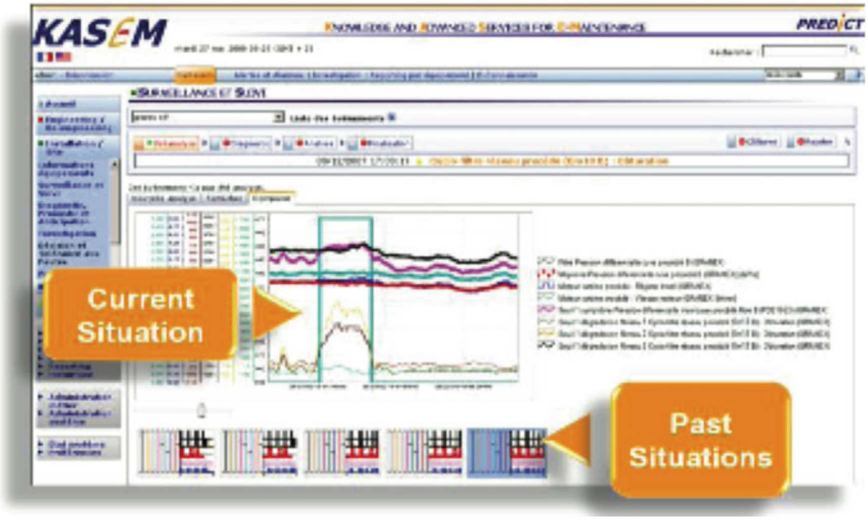


Fig. 28.4 Example of fleet-wide diagnosis facilities

28.4 Conclusion

In this paper, a global methodological approach which is applied throughout the life cycle of a fleet is proposed. It allows to design and support proactive fleet health management. The underlying methodology incorporates different engineering methods and encompasses the expert knowledge and skills on equipment and system by including integrated technologies and tools within the KASEM software platform. It enables a consistent fleet-wide health management by ensuring a consistent information treatment process from monitored data to fleet-level indicators. Furthermore, information and knowledge sharing is enabled by means of semantic modeling within the fleet dimension.

As a result, this approach provides user information from different service areas: Production/Maintenance; at different responsibility levels: Technician/Engineer/Manager with a comprehensive and effective proactive decision management support regarding both the global health of the fleet and the present status of their maintenance efforts.

References

1. Iung B, Levrat E, Crespo Marquez A, Erbe H (2009) Conceptual framework for e-maintenance : illustration by e-maintenance technologies and platforms. Ann Rev Control 33(2):220–229
2. Verma AK, Srividya A, Ramesh P (2010) A systemic approach to integrated E-maintenance of large engineering plants. Int J Autom Comput 7(2):173–179

3. Bonissone PP, Varma A (2006) Predicting the best unit within a fleet: prognostic capabilities enabled by peer learning, fuzzy similarity, and evolutionary design process. J Comput Stat Data Anal 51(1):398–416
4. Monnin M, Leger JB, Morel D (2011) Proactive facility fleet/plant monitoring and management. In: Proceedings of 24th international congress on condition monitoring and diagnostics engineering management. Stavanger, Norway
5. Léger J-B (2004) A case study of remote diagnosis and e-maintenance information system. In: Keynote speech of international conference on intelligent maintenance systems. Arles, France
6. Peysson F, Ouladsine M, Outbib R, Leger JB, Myx O, Allemand C (2008) Damage trajectory analysis based prognostic. In: Proceedings of IEEE international conference on prognostics and health management, pp 1–8

Chapter 29
Managing Risks in Service Value Networks

T. Uusitalo, K. Palomäki and E. Kupi

Abstract Development of successful service concepts is a prerequisite for success. Developing business through service excellence is one way of surviving competition while improving customer service and loyalty. Service value is created in a network context, and the structure and dynamics of the value network and customer expectations influence the complexity of service delivery. This needs to be taken into account when considering the management of risks. The traditional emphasis on of risk management has been on protecting the system, and its users, from the failures in the system. When considering the performance of a system in its larger commercial and political environment, uncertainty may provide opportunities as well as threats. The key research question in this paper is how to develop new service business and manage risks in service value networks taking into account both threats and opportunities related to the services' value creation.

29.1 Introduction

Moving towards service-oriented business models introduces new risks that must be taken into account in the development of new services. There is a need to consider risks in customer collaboration, information exchange among service value network partners and capabilities related to collaboration and communication. Internal risks are related to capabilities required by the high-intensity customer relationships of the new services. External risks include e.g. how customers are willing to share critical information needed for successful delivery of new services [1].

T. Uusitalo (✉) · K. Palomäki · E. Kupi
VTT Technical Research Centre of Finland, 1300, 33101 Tampere, Finland
e-mail: teuvo.uusitalo@vtt.fi

E. Kupi
e-mail: eija.kupi@vtt.fi

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_29,

Service value is created in a network context; the complexity of service delivery is influenced by the structure and dynamics of the value network and by customer expectations. This needs to be taken into account when considering the management of risks. Value for customers is created at the network level, each network partner contributing incremental value to the overall offering. Network actors contribute to the value creation process by focusing on their core competence and cooperating with other network actors—such as suppliers, partners, allies and customers—through various value constellations [2].

ISO/EIC Guide 73 defines risk as the "effect of uncertainty on objectives". The effect is a deviation from the expected and can be positive or negative [3]. Risk management is about improving business performance via systematic identification, appraisal and management of business-related risk. In addition to downside risks, improving performance also covers exploiting opportunities or favourable possibilities—upside risks. In service business management, it is important to take risks into careful consideration e.g. during the transition process of an organisation from manufacturer to service provider [4].

The aim of this paper is to present results from a study which examined risk management in service business networks. Firstly, and based on existing literature, the paper discusses risks and success factors in new service business development and service business networks. This is followed by presentation of case study results regarding the management of risks in service business networks and finally by discussion on the significance of the results.

29.2 Risk Management in Value Networks

New networked business models challenge the traditional theory of the firm. The new business models are based on complicated inter-organisational systems for innovation, production, marketing and connecting with customers and investors. On the supply side, outsourcing and the Internet have led to new connections that provide a variety of services. On the demand side, open architectures and empowered consumers are driving both innovation and customisation of existing products and services and their marketing. This requires companies to view their profits and risks in terms of not what they control internally but their capabilities related to the networks in which they are embedded [5].

While networks present opportunities, they also raise a number of risks and challenges. Accountability and transparency can be lost in the complex web of networks. Business networks are increasingly complex and dynamically changing webs of relationships. Key drivers are increasing product/service complexity, e-business, outsourcing and globalisation [6]. Examples of network threats include terrorism, disease and global warming. These risks are the downside of the opportunities reflected by the new value networks. New networks have provided considerable increases in access to resources and markets. At the same time this change has exposed companies and economies to new complexities and risks because of the increased interdependencies implied by the networks. Management needs to navigate and balance the trade-offs between these opportunities and risks [5].

Risk management methodologies have been developed as the appropriate unit of analysis for an individual organisation. Lately, the focus has been shifting towards the extended organisation and the network level. As companies become increasingly reliant on different collaborative arrangements in their business models they also become more dependent on the capabilities and resources of other companies. This makes their situation more unpredictable regarding possible changes in the business environment [7].

The current models and frameworks for enterprise risk management (ERM) emphasise the need to embed risk management systems within business processes. The focus is usually concentrated on the individual organisation as the appropriate unit of analysis, bearing little or no relation to the company's involvement in wider networks. It has been argued that the recent financial crisis is largely the result of a failure to represent and understand interconnectedness in this wider sense of embeddedness in networks [8].

The general risk management framework for supplier networks consists of risk identification, risk assessment, decision and implementation of risk management actions and risk monitoring. Risk identification is an essential part of the risk management process because risks in networks may arise from complex chains that are difficult to perceive [9]. The dynamics of relationship development causes an additional difficulty. Companies in networks should therefore communicate and share their views on risks, because different views help in recognising and understanding common opportunities and threats. There is a need for closer integration between partners in the primary level supply chain [10]. The improved understanding helps make for better decisions and decreases the risks for both a single organisation and an entire network.

Hallikas and Varis [7] review different approaches to risk management in value networks and present different theoretical backgrounds for analysing networks. They conclude that the network key actor has an important role in maintaining the health of the whole business ecosystem. This implies that the risk management capability of key actors should be directed towards orchestration of the entire network. The key actors should filter the changes in business environment and customer needs to other actors, as well as provide new value-creating opportunities for connected actors. The challenge for risk management lies in identifying the trends and uncertainties that may change in the value network, and in the degree of interdependence among actors from the point of view of risks in the network.

Networks increase interdependencies, and this creates challenges for managing risks. This is especially apparent in areas such as security and enterprise risk management, where the actions of a single player can cause problems for everyone in the network [11]. The networks have led to second-order and third-order effects that are absent from the bilateral relationship. Networked enterprises raise issues of group influence, cascading, contagion and interdependent risks that cannot be controlled through standard mechanisms [5].

29.3 Risks and Success Factors in New Service Business Development

In order to manage risks systematically, organisations are typically applying formal risk management methods and procedures. In addition to formal practices, organisations need to increase flexibility to absorb the unexpected and to be prepared for the uncertainties that cannot be assessed beforehand [12]. The understanding of risk and uncertainty management has mainly been based on linear models that describe how various risk events are present in considered business activities. These models provide valuable information as to why firms should manage risks. They provide less information, however, on how firms turn the incentives to manage risk into actual decisions on the choice of risk management measures, on how risks change, or on how the risk management strategies evolve over time [13].

Uncertainty can lead to opportunities and success, but also to threats that at worst can end in failures. Factors that positively affect development of new service business are critical if an organisation wants to succeed. If, however, the organisation ignores or falls down on these factors, there is a risk that service development may fail. Risks can be considered as the other side of the coin to success factors [4].

Different risks and success factors are critical at the different stages of a new service development process. Johnson et al. [14] introduce a process cycle of new service development. In the development stage of a new service, new ideas are screened, and winning concepts developed and tested for feasibility. The first steps of the new service development process require deep understanding of the customers' behaviour and needs, production and business processes. Concepts that pass the development hurdle are then considered in the analysis stage, to determine their potential as part of a profitable business venture. At this stage, it is important to consider the value network structure and information flows. Successful concepts then move to the design phase, where a new service product and process is created and field tested with appropriate personnel training and marketing campaign. Finally, a proven new service is given a full launch. The new service development process is driven by enablers: teams that are cross-functional, tools and an organisation context that includes a culture of accepting innovation. A service product consists of people, systems and technology.

In service companies, the foundation for quick and effective service development lies in the strategic operation choices made regarding the use of teams, information technology and the process design of new service development [15]. In addition to these, success factors influencing new service development include e.g. management support and involvement of both managers and other employees, interaction with customers, a skilful project leader and a formal new service development process that also has some elements of improvisation. Besides external communication, managing internal communication among the company's personnel is also important. [e.g. 16–19].

The role of customers has been recognised as especially important, and the more successful new service developers have typically made efficient use of information about the customer in idea generation, business evaluation and marketing plan preparation [19]. Understanding the customer is the key to commercial success; attention should therefore be paid to the customers' needs, expectations, quality perceptions and values when developing and launching new services. The service culture and strategy of a company should be customer-centric and in line with customers' values and priorities. Involving the customer in the actual development process is also important. However, besides customers, the company needs to monitor and understand market and future trends [17]. Development of new services should be based on an intimate knowledge of customer needs, problems and operating systems. Also critical for the success of new services are involvement in the process of expert front-line personnel and the implementation of a formal and planned launch programme [20].

In addition to the importance of customer orientation in service business development, there are also great challenges and risks in the collaboration between customer and supplier, as well as in the activities of the service network [4]. Possible obstacles may relate to lack of information on specifications, goals and markets, uncertainty about the sponsor, difficulty in dividing responsibilities and in resource allocation and a lack of systematic documentation, reporting and feedback [16]. Besides these obstacles and challenges, the aforementioned success factors must also be considered in order to allow better preparation for making use of opportunities, on the one hand, and on the other to avoid the pitfalls and risks when developing new service business.

29.4 Case Study

29.4.1 Case Study Method

The methods used in this study involved a literature review and a multiple case study. The literature review dealt with existing literature on how risk management should be integrated in service value networks and service business development. The case study is the preferred strategy when "how" or "why" questions are being posed, when the investigator has little control over events, when the focus is on a contemporary phenomenon within some real-life context, and when new areas of research are explored [21, 22].

The case study was carried out in four companies, two from the manufacturing sector and two from the real estate business. Specific service business cases were chosen from each company for the purpose of understanding the actual value networks and risk management processes in service business. The service business cases from the manufacturing companies were delivery of after-sales services, and service business as part of new product offering. The real estate cases were related to the development of rental services. A key step in this case study involved within-case analysis, with detailed case study reports for each site. These reports are central to the generation of insight because they help researchers cope early in the analysis process with the volume of data. This process allows the unique patterns of each case to emerge before investigators push to generalise patterns across cases [23].

The data gathering in the case study included interviews, working groups and workshops. The key personnel within the companies were first interviewed to establish the current status and the key development areas. The interviews were recorded and transcribed. The data of the initial interviews were used for familiarisation with the company and its current practices and key development targets, and analysed to create an overview of the main challenges and strengths related to the value network and new service business development.

The next step was to establish a working group involving the company representatives. A series of workshops were organised by the group, where threats and opportunities related to the new service business were discussed. Various methods were used, including market study, PESTE analysis, business network analysis and roadmapping.

29.4.2 Case Study Results

The analysis of the case study data showed that the companies had set targets for developing new product offering and related new services. It was not clear, however, how these targets could be reached. The market potential had been analysed, but there was a need to obtain more detailed information on the market potential of the service business, and on international market opportunities in particular.

The study of the threats and opportunities related to new services revealed that there were significant external and internal factors that the company needed to take into account when developing new services. Threats were related to reaching new customers, collaboration with new partners, failure to develop new product and service offering, lack of required resources and competence and new competitors. The new services can provide opportunities in expanding the customer base, increasing revenue and forming new profitable partnerships and long-term relationships with customers.

The case companies' service value network risks were related to suppliers and supplier networks, clients, stakeholders and competitors. Risks related to suppliers included failure in the selection of partners, lack of competence, lack of resources, lack of communication and poor financial status. Risks related to the supplier network were similar to the supplier-related risks, but in the absence of direct contact with the suppliers' suppliers were more difficult to control. Stakeholder-related risks were related to e.g. authorities' requirements, delays in various permit procedures and communication problems. Clients' requirements and understanding of client needs were the common risks concerning clients. The risk factors related to the value network are detailed in Table 29.1. The risk factors can be seen to include both threats and opportunities.

The results of the case study identified the following key challenges concerning the management of risks in the development of new service business:

- Allocation of resources for the development of new service business.
- Networking and forming partnerships required for delivering new services.
- Designing the new service business offering.
- Concurrent development of new technologies and related services.
- Setting of clear objectives for the development of new services.
- Following changes in the operational environment.

Based on the results of the case study, the following key steps in managing risks related to new service business development were identified: (1) analysis of market potential of the new services, (2) analysis of risks related to internal resources and capabilities available for the service development, (3) analysis of external risk factors affecting the business environment, (4) identification of the key players and their role and related risks in service business network, and (5) development of measures that mitigate threats and utilising the opportunities related to new service business. When the services have been launched and are on the market it is important to monitor the development of risks. Risk monitoring must cover the risk factors related to the value network.

Table 29.1 Risk factors related to the value network of the case companies

Value network actor	Risk factors
Company's internal organisation	Competence
	Resources
	Coordination of network activities
	Communication
Supplier	Selection of partners
	Competence and resources
	Communication
	Financial status
Supplier's network	Scheduling of activities
	Visibility to the suppliers' suppliers
	Communication
Stakeholders	Requirements
	Permit procedures
	Communication
Clients	Understanding of client needs
	Communication

29.5 Discussion

The aim of this paper was to present results from a study which examined risk management in service business networks. Based on existing literature, the paper discussed the key factors related to risk management in value networks, as well as risks and success factors when developing new service businesses. The case study results indicate that service businesses change business and the risk environment, forcing organisations to identify new risks and to adapt their thinking about risks and the ways of managing risks. The results of the case study show the importance of exploring both the internal and external environments of an organisation, and their future development, when analysing new services and the related threats and opportunities. There is a need for risk assessment and management to cover the entire service business value network.

The identified risks related to value networks concern network coordination, control and communication. This study can be the basis for proposing a framework for integrating risk management in service value networks. The framework includes five main phases: market potential analysis, internal resources and capability analysis, external business environment analysis, value network risk analysis and development of risk management measures. Risks should be identified in each of these phases, their significance assessed, and risk management measures designed and implemented. Risk monitoring must also be organised to cover the risk factors in value networks.

Restricting analysis to four in depth case studies has placed limitations on this research, as the significance of the results with regard to the broader organisational population can be questioned. Further research involving practical application of the approach proposed in this paper would therefore be necessary.

References

1. Gebauer H, Fleisch E, Friedli T (2005) Overcoming the service paradox in manufacturing companies. Eur Manage J 23(1):14–26
2. Basole RC, Rouse WB (2008) Complexity of service value networks: Conceptualization and empirical investigation. IBM Syst J 47(1):53–70
3. ISO/IEC (2009) Guide 73 risk management—vocabulary. International Organisation for Standardisation
4. Ojanen V, Lanne M, Ahonen T, Tuominen M (2008) The customer-centric development of new industrial services: antecedents, risks and their management. In: Proceedings of the 1st ISPIM innovation symposium
5. Kleindorfer P, Wind Y (2009) The network imperative: community or contagion? The network challenge. strategy, profit and risk in an interlinked world. Pearson Prentice Hall, NJ, pp 3–24
6. Harland C, Brenchley R, Walker H (2003) Risk in supply networks. J Purchasing Supply Manage 9(2):51–62
7. Hallikas J, Varis J (2009) Risk management in value networks. Supply Chain Risk: Int Series Oper Res Manage Sci 124(1):35–52

8. Power M (2009) The risk management of nothing. Acc Organ Soc 34(6–7):849–855
9. Hallikas J, Karvonen I, Pulkkinen U, Virolainen VM, Tuominen M (2004) Risk management processes in supplier networks. Int J Prod Econ 90(1):47–58
10. Ritchie B, Brindley C (2007) An emergent framework for supply chain risk management and performance measurement. J Oper Res Soc 58(11):1398–1411
11. Kunreuther H (2009) The weakest link: risk management strategies for dealing with interdependencies. The network challenge. strategy, profit and risk in an interlinked world. Pearson Education, NJ, pp 383–398
12. Williamson D (2007) The COSO ERM framework: a critique for systems theory of management control. Int J Risk Assess Manage 7(8):1089–1119
13. Lanne M (2008) The role of risk management in industrial service business. In Vähä P, Salkari I, Alahuhta P, Leviäkangas P (eds) VTT symposium on service science, technology and business, pp 63–72
14. Johnson SP, Menor LJ, Roth AV, Chase RB (2000) A critical evaluation of the new service development process: integrating service innovation and service design. In: Fitzsimmons JA, Fitzsimmons MJ (eds) New service development—creating memorable experiences. Sage Publications, CA, pp 1–32
15. Froehle CM, Roth AV, Chase RB, Voss CA (2000) Antecedents of new service development effectiveness—an exploratory examination of strategic operations choices. J Serv Res 3(1):3–17
16. Edvardsson B, Haglund L, Mattson J (1995) Analysis, planning, improvisation and control in the development of new services. Int J Serv Ind Manage 6(2):24–35
17. Edvardsson B, Gustafsson A, Enquist B (2007) Success factors in new service development and value creation through services. In: Spath D, Fähnrich KP (eds) Advances in service innovations. Springer, Berlin, pp 165–183
18. Johne A, Storey C (1998) New service development: a review of the literature and annotated bibliography. Eur J Mark 32(3–4):184–251
19. Martin CR, Horne DA (1995) Level of success inputs for service innovations in the same firm. Int J Serv Ind Manage 6(4):40–56
20. De Brentani U (2001) Innovative versus incremental new business services: different keys for achieving success. J Prod Innovation Manage 18(3):169–187
21. Yin RK (2008) Case study research: design and methods, 3rd edn. Sage Publications, California
22. Eisenhardt KM (1991) Better stories and better constructs: the case for rigor and comparative logic. Acad Manag Rev 16(3):620–627
23. Eisenhardt KM (1989) Building theories from case study research. Acad Manag Rev 14(4):532–550

Chapter 30
Towards an Asset Management Reference Model: Basis for a Unified Approach

Y. Wijnia, J. de Croon and J. P. Liyanage

Abstract Asset Management is a concept that has a very wide range of uses and different levels of maturity across diverse industrial sectors and regions. Because of this, there is lots exists a large amount of variation in the concepts and the terminology used. In discussing the concept of asset management, the asset management professionals tend to adhere to their own definitions. This situation is an impediment to bringing the concept further, and can mostly be attributed to the fact that framing an improvement in one specific approach or terminology may render it virtually inapplicable in another context. In other words, quite comparable approaches and concepts have to be invented over and over again. To prevent the chaos and in order to provide the basis for a unified resolution, a system level reference model for translating improvements into other realms would be very helpful. In this paper, the outline of such a reference model is developed.

30.1 Introduction

Since the 1990s, the term Asset Management has seen a significant increase in use [1]. If one were to observe a live discussion or a debate on Asset Management, odds are that the financial news are analyzed or put in perspective by someone

Y. Wijnia (✉) · J. de Croon
Asset Resolutions B.V., PO box 30113, 8003 CC Zwolle, The Netherlands
e-mail: ype.wijnia@assetresolutions.nl

J. de Croon
e-mail: john.de.croon@assetresolutions.nl

Y. Wijnia
Faculty of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands

J. P. Liyanage
Faculty of Science and Technology, University of Stavanger, Stavanger, Norway
e-mail: j.p.liyanage@uis.no

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_30,

from the Asset Management field representing or focusing on the performance of any big financial institute (ABN AMRO, RBS, HSBC, to name a few). In the credit crunch of 2008–2009, debate was often on toxic or high financial risk assets and in normal operations terminology, such as shoulders, head, the neckline, perfect storm, cup and handle, resistance line, among others. Such a discussion is completely incomprehensible if one is from outside the trade or the profession. Yet, working through the concepts that are behind the labels, most often the development of prices or options for traded securities are involved, and the discussions in the end concentrate on whether to buy or sell certain 'financial assets'.

Another example on the more technical side is from the field of maintenance. About 25 years ago, maintenance was about the four principals of maintaining technical equipment: cleaning, aligning, balancing and lubricating, or rather retaining physical equipment in an acceptable technical condition. Nowadays, it is often seen covered with the term of Asset Management , asset lifecycle management, or engineering Asset Management . In essence, Asset Management is still seen based on the same four basic actions, and nothing fully realizable of this down to earth approach and timely relevant concept can be found in any definition of Asset Management . See for example the definition used in the PAS55 [2]:

> Systematic and coordinated activities and practices through which an organization optimally manages its physical assets, and their associated performances, risks and expenditures over their lifecycle for the purpose of achieving its organizational strategic plan.

Not wanting to ridicule the effort that has been put in to arrive at some common understanding on the challenging things that you need to do to keep your equipment fit for purpose, it should not be surprising that this definition does not get across the operational and technical maintenance engineers. In their opinion, this may just be another bunch of big, fancy words of someone who does not know how to change a light bulb. It also underlines that there is a significant communication as well as knowledge gap between the technical and the business worlds, resulting in very abstract ideologies in resolving the matters that are of true importance to the industrial sectors. Despite the hype on Asset Management , the relationship between more technical and managerial issues still remains immature.

A third view to be covered here is that of the energy sector, for instance energy distribution network operators of the Netherlands. When the concept was introduced around 1999 (not coincidently the year of market liberalization), it had for many novice asset managers the connotation of postponing investment. Given the context of a very high performance of the grid [3, 4][1] and a very low expenditure on maintenance [5][2] this is perfectly understandable. Investments in quality and capacity are the most significant ones at the discretion of the company, so there are very little other options. Nevertheless, looking at the situation very closely, it is

[1] The Dutch grid for distribution of gas and electricity ranks among the most reliable and safest grids in the world.

[2] Enexis reported 20 million euros of preventive maintenance on an asset base with a book value of 5 billion euro, which equals 0.4 %. If replacement value is used, it is below 0.2 %.

significantly different from the concept as defined by existing reference frames, for example PAS 55 and alike.

Looking at the three cases that were briefly noted above, it is obvious that the governing terminology as well as the conceptual resolutions are quite different from one another. Drop a maintenance engineer on the trading floor and there will be nothing physical to touch: the assets discussed are essentially claims to some (future) cash flow of other assets. Ultimately, there are objects in the real world behind the assets, but there may be several layers of abstraction between reality and the assets traded. On the other hand, employ a financial asset manager in an infrastructure context, and he might be stunned by the idea that investments are one way traffic. You can put your money in, but getting it out by selling the asset for infrastructures is quite impossible, especially if they are buried like cables and pipelines. Finally, if the infrastructure asset manager has to work in a maintenance environment, for example a plant full of rotating equipment, none of the familiar tools will work. It is not just putting an asset in leave it for 10 years or more like in infrastructures, assets in a plant require continuous attention, either in the form of operation or maintenance. The work of the plant asset manager does not stop at the investment decision, it just starts there.

Yet, despite the obvious differences, there seems to be a deeper unexplored commonality between the fields. In principle, every asset manager is in his or her own way and context trying to optimize value, and it constitute a functional definition per se. Thus, the many application fields, including the ones discussed here, have the potential to derive effective solutions from a single generic concept. If that is the case, lessons may well be learned between the fields contributing to value creation process based on know-how across the fields as well as sectors. However, for this to be possible, as well as pragmatically usable, the generic concept should not be too far out. If the cash flow is the only idea shared between the fields, there is not much to learn. Given that Asset Management captures a far more holistic process of value creation that has both short term and long term implications, at least some kind of common framework is needed. In this paper, the idea of such a generic framework will be further explored.

30.2 The Analytical Framework

Developing an analytical framework to compare different fields in order to arrive at a generic framework sounds like the Munchhausen trick of pulling yourself out of the swamp by your own bootstraps. Theoretically, one needs the generic framework to identify the areas for which the different applications should be compared to see how far they have drifted apart. However, for all practical purposes, a bit of common sense can also be good enough to get things started. For example, Komonen et al. [6] have shown that in engineering Asset Management there is a relation between the asset characteristics, the market in which the asset is operated and the optimal asset strategy.

The idea behind this framework can be found in the recognition that the length of the planning period affects the level of uncertainty involved in the prognosis and thus the risk involved in business decisions. The longer the planning period the more uncertain predictions become. Stacey [7] divided the future development into three environments. The first is the Stable region, where the business predictions are reasonably certain, with uncertainty measured in percentages. This is also referred to as a closed future. The second environment is the Probabilistic environment. Predictions are uncertain, but there is some knowledge on the amount of uncertainty, and order of magnitude estimates tend to be correct. This is also known as the contained future. The third environment is volatile. Here predictions are fundamentally uncertain, and may be orders of magnitude wrong. This is the open ended future. Drivers behind this volatility can be diverse, but tend to be in new product development, changing market needs and changes in production technologies. On short term, market needs tend to be stable, but new product introductions can result in substitutions, even without changing the fundamental needs. The introduction of digital cameras for example did not alter the "need" to make pictures, but it almost eradicated the market for films. With regard to production technology, an example can be found in the transition in propulsion technology in shipping. The need for propulsion did not fundamentally alter, yet diesel engines economically outperformed steam engines by far and thus many steam driven ships were scrapped.

Relating this recognition to the asset strategy, if the (technical) asset life is comparable to the lifecycle of the product it produces, decisions with regard to the asset are in the stable future environment. The asset strategy then should focus on keeping the asset fully operational during its lifecycle. But if the asset life is short compared to the market, it might be interesting to develop more durable assets. On the other hand, if the assets life is much longer than the product lifecycle, it might be wise to invest in flexibility of the asset, so that the asset can be used for something else once the demand dries out. The market and technological environment thus drive the high level asset strategy and therefore the fundamental questions in Asset Management .

Another option is to use the asset lifecycle as the organizing principle for the analytical framework. The observation behind this idea is that asset managers in the gas and electricity distribution infrastructure focus on investment decision (the design phase), whereas asset managers in production facilities focus on operations and maintenance decisions (the operational phase). Given the characteristics of the system in which the asset managers operate, this is not more than logical. The infrastructure assets are mostly cables and pipelines, which are difficult to access as they are buried. The only option is remote diagnostics like leak finding programs and partial discharge measurements, but their predictive power is not very impressive. Fortunately, there is virtually no wear and tear in their use, so they do not require much maintenance. Furthermore, they are passive elements, no action is needed to make gas and electricity flow through the assets. In a production facility, operation is everything. If no raw materials are fed into the plant, nothing will come out. Optimization of the operational parameters can result in a much better output of the plant. In chemical plants, for example, the improved understanding of the process and de-bottlenecking of the plant typically results in the end of life capacity being factors higher than the design capacity. And the wear and tear of the (rotating) equipment makes maintenance a necessity.

Many similar examples can be used to show that there is a link between the asset characteristics and the main focus of Asset Management . The idea was explored in the EURENSEAM meeting in Seville 2009. During that session, the following list of potentially relevant asset characteristics was produced:

1. Function: Commercial versus public
2. Time Scale: Short (weeks) versus long (50 year +)
3. Location: Fixed versus moving and Point versus distributed
4. Economic/technical assets: relative importance of assets
5. Environment (external effects)
6. Technology (mechanical, civil, electrical, software)
7. Timescale: split into Life span and reaction time
8. Market behavior: stable versus dynamic
9. Technology development: fast versus slow

Items 8 and 9 on this list in fact are the business environment parameters of the framework in Fig. 30.1. Such attributes can effectively be used to identify the business framework that can be used for analyzing an organization's Asset Management system.

To test whether these characteristics were related to the way Asset Management was employed, a series of interviews was conducted with Dutch infrastructure asset managers [8]. The framework used in those interviews is presented in Fig. 30.2. The framework helped in getting a quick overview of the asset system.

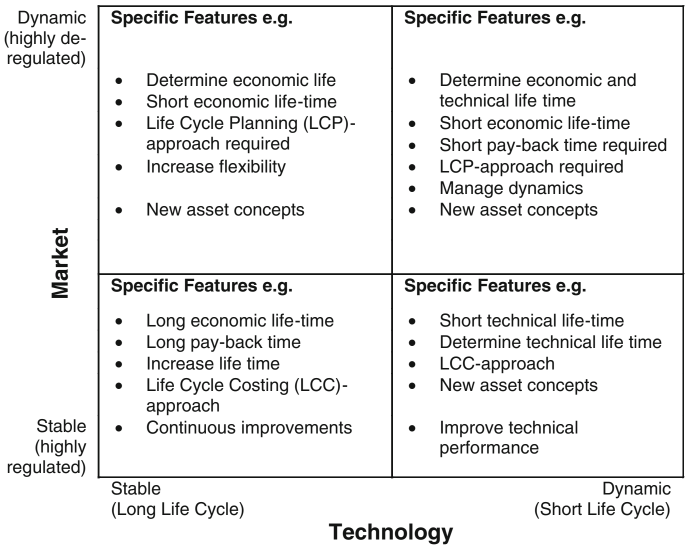


Fig. 30.1 The influence of various business environments on asset strategies [7]

Asset characteristics • Function • Behavior • Life cycle length • Location • Type	**Environment** • Market • Technology development
Key life cycle phase • Key life cycle phase	**Management attention** • People • Key performance indicators • Budgets • Processes • Systems

Fig. 30.2 The framework for analyzing Asset Management systems

An important result of those interviews was that the interviewed organizations had a strong tendency toward specific lifecycle phases. This to a certain degree can be attributed to setting the decision basis depending on where the more dominant business case exists with respect to an organization's strategic position and industry potential. This is in contrast to the widely advocated integrated lifecycle approaches that focus on a seamless interface along the lifecycle. Today this largely remains a 'faith' than a fully justified or validated case. There was not a clear result in the management system to head up toward a long lasting resolution. This may have been affected by the volatile markets, financial uncertainties, and other hidden vulnerabilities of doing business. Asset owners tend to be keen on retaining additional business capacity to cope with hidden risks better than dedicating that to a long-term course. The situation can also be seen affected by the organizational history or legal constraints to a certain extent.

Given that this framework helped in differentiating the Asset Management systems of infrastructure managers, it will be used at a later point in time as a starting point for the wider analysis across other industrial sectors.

30.3 Common Ground

Despite the impossibility of using the same management instruments for all assets, in terms of the process there is common ground as can be seen in Fig. 30.3. For every asset as explored in this paper, the central function of the asset is to deliver value. It depends on the context what that value is. In private settings generating income is generally more important, whereas in the public context other values like accessibility and equality may be more relevant. Yet, the assets have a purpose, not an intrinsic value. But purpose is something subjective, something the asset owner is free to define. Starting point of any form of Asset Management thus

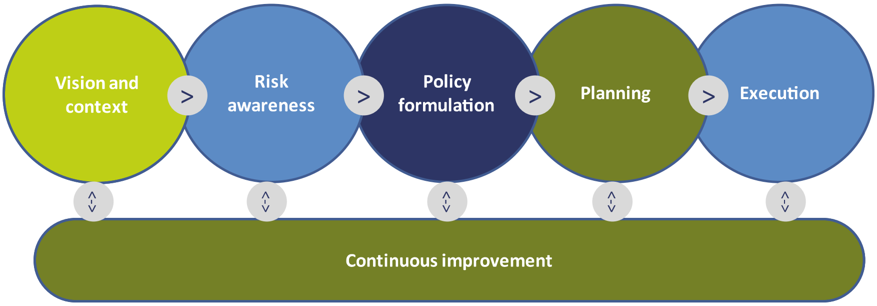


Fig. 30.3 The generalized Asset Management process

seems to be framing the context and formulating a vision and mission for the context in which the company operates.

Furthermore, the risks of the assets seem to be an explaining factor for the management attention. If the value of the assets depends on market emotions, monitoring the market to buy and sell assets at the right time is the right thing to do. Only with a long term perspective one could rely on the fair value of the underlying assets. Likewise, if a breakdown of equipment stops the facility producing goods and thus stops the generation of income, it is not surprising that Production facility asset managers focus on maintenance and operation. Similarly, disruptions and congestion of infrastructures create a significant societal cost, so it is perfectly understandable that infrastructure asset managers focus on providing capacity beyond what is economically feasible from the company perspective. It seems therefore fair to state that risk awareness is the second stage of any form of Asset Management . For all types of assets, risks need to be mitigated if these risks are not acceptable. Mitigation measures are required with a different time horizon as can be seen in Fig. 30.1. For the longer term policies need to be defined to reduce non acceptable risks like to increase flexibility for the stable assets in a deregulated market. In the planning process, the defined policy is made specific. e.g., for utilities companies specific specifications could be set up to increase cable capacity in case a failure has occurred and the cable needs to be replaced.

Finally, executing the defined mitigation measure is essential for any form of management. It will not be discussed in this paper. However, the continuous improvement of performance seems to be of major importance for Asset Management . Only by learning from the past the incredible levels of performance as witnessed today are possible.

30.4 Conclusion

Undoubtedly, the discipline of Asset Management has shown a clear growth of interest over the last few years. Despite the developments this discipline has not yet managed to provide a clear and a unified basis that can provide a good reference across different industrial sectors. This paper presented some early work to address this issue, and brought some insights from the energy distribution and other sectors where convincing practical matters exist showcasing the difficulties associated with the use of existing terminology and concepts. A unified approach can bring many benefits to the different user groups inclusive of a common reference and for further development of this important discipline.

Acknowledgments Many insights were developed in the meetings with the Eurenseam Network in Sevilla (2009) and Delft (2011). Ideas have been tested in an interview round conducted in the Netherlands in 2010. The authors wish to thank EURENSEAM members and the interviewed asset managers for their contribution.

References

1. Burns P (2010) The history project. Available from: https://www.amqi.com/index.php/historyproject. Cited 24 June 2011
2. British Standards Institution (2008) PAS 55 part 1: specification for the optimised management of physical assets. British Standards Institution, UK
3. Netbeheer Nederland (2010) Betrouwbaarheid van elektricteitsnetten in Nederland 2009. Arnhem
4. Netbeheer Nederland (2009) Storingsrapportage gasdistributienetten 2008. Arnhem
5. Alkema B (2011) Asset Management Proces: Investeringsbeslissingen en onderhoud. Asset Lifecycle management, Rotterdam
6. Komonen K et al (2010) Investments, capacity and maintenance: ways to safely increase capital turnover. In: Proceedings of the Euro maintenance international conference and exhibition. Verona
7. Stacey RD (1990) Dynamic strategic management for the 1990s—balancing opportunism and business planning. Kogan Page Ltd, London
8. Van der Lei T, Wijnia YC, Herder PM (2010) Towards an integral Asset Management framework using asset characteristics, asset environment and lifecycle phases as leading principles. In: Proceedings of the 5th World conference on engineering asset management 2010. Springer, Brisbane

Chapter 31
Performance Standardization for Sustainability in Complex Production Networks: A Roadmap

J. E. Beer and J. P. Liyanage

Abstract Sustainability has attracted the attention of many researchers and organizations. There is a Europe-wide interest in measuring sustainability performance. A standardized framework to determine sustainability performance of dispersed production clusters, however, has yet to be found. Existing initiatives either focus on partial aspects of sustainability or do not support sustainability considerations for complex value networks but rather focus on the organizational level. This paper discusses some problems related to existing initiatives and introduces *SustainValue*, a project that responds to the known shortcomings and that aims at the creation of a governing framework for sustainability performance.

31.1 Introduction

The numeric pressure due to human population growth, changed demographic structure, and changed consumption patterns on a fixed base of natural resources [1] has made sustainability a key topic in academic and industrial debates. Given the growth of production clusters around the globe involving many agents in business transactions, reconsidering the basic assumptions of how business is done in Europe may be necessary to ensure as well as to enhance the business advantage, i.e., new business models have to be evaluated with respect to economical as well as environmental and social aspects of sustainability. Besides business models, logistical systems that proved to be beneficial in economic terms such as

J. E. Beer (✉) · J. P. Liyanage
University of Stavanger, 4036 Stavanger, Norway
e-mail: jakob.e.beer@uis.no

J. P. Liyanage
e-mail: j.p.liyanage@uis.no

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_31,

Just-in-Time might need to be assessed with respect to their environmental and social sustainability [2]. What is missing, though, is a set of tools and methods by which businesses can assess, simulate, and design new business models and performance in complex value networks [3]. This relies on the claim that traditional transnational manufacturing networks, solely built on comparative cost advantages, cannot sustain the natural and social environment to the extent it is necessary. This is the focus of the newly established European project on ‘Sustainable value creation in manufacturing networks (SustainValue)’ that aims at making a meaningful contribution to this effort.

The purpose of this paper is to make the case for *SustainValue,* a project that will explore a new framework for sustainability performance standardization, as a response to call NMP.2010.3.1-1 of the European Commissions’ Seventh Framework Programme [4]. The entire SustainValue project in general addresses principal requirements of the current business environment by developing industrial models, solutions and performance standards for new sustainable and more performing value-adding networks.

31.2 Status Quo Review

31.2.1 Growth of Production Clusters and Networks

With many companies taking the opportunity to expand their supplier base, their production, and their customer base globally, the way companies perform has changed significantly with respect to scale, scope, and requirements on business processes. In the course of this process, stakeholders to consider and activities to be managed and monitored have changed [5]. Cheap labor and other production inputs do not represent decisive competitive advantage anymore in many cases as they can be accessed at any time in many markets [6]. End products are directed to global markets that often times represent a bigger share of customers and contribute more to company or cluster growth than local markets do. For instance, China is a more important market for the Mercedes-Benz S-Class than Germany or the US [7]. Outsourcing of vital business operations, sometimes even core competencies, has become quite common, which can be understood with regard to the complexity of certain industries (e.g., the automobile industry). Driven by increasing product complexity and lower barriers for international trade and cooperation, enterprise-led value creation has transformed into enterprise-centered production clusters over time. Eventually, organizations that were part of these clusters have diversified their products and services and transformed from dedication to production clusters within clearly defined boundaries into network partners, offering value creation to multiple complex clusters. As a result, the number of entities involved in business processes has grown beyond the traditional levels.

31.2.2 Establishment of Dynamic and Complex Value Chains

Traditionally, many businesses followed patterns that could be adequately described with models such as Porter's Value Chain Concept, evolving around the idea that an organization conducts a set of primary and support activities that creates value customers are willing to pay for. In the Value Chain theory, differences among competitor value chains are considered "a key source for competitive advantage" [8]. With value chain primary activities being inbound logistics, operations, outbound logistics, marketing and sales, and service, this concept is hardly applicable to the sources of competitive advantage in many of today's industries. In the automobile industry, for instance, a great portion of value added to the end product comes from the OEMs' wide-spanning supplier networks (as opposed to supply chains—a distinction Urban [9] specifies), and competitive advantage or disadvantage hence to a good extent is determined before the OEM even chimes in. Competition in general will be and already is about how well networks rather than individual organizations perform [9]. Similar conclusions about the applicability to today's manufacturing environment can be drawn with respect to other theories such as the Resource-based View theory which concentrates on an organization's internal resources as source for competitive advantage [10] and hence do not provide adequate depiction of the complex and comprehensive value networks and production clusters that have replaced traditional business models in many industries. On the other hand, Stakeholder theory could possibly gain much more significance for the understanding of business as the number of stakeholders (i.e., entities "who count" [11]) directly involved in value creation has grown compared to traditional value chains. From a shareholder perspective, the value of a company is determined based upon the performance it shows in meeting the shareholder expectations [12], which increasingly often goes hand in hand with sustainability performance.

31.2.3 Frameworks and Methods to Track Performance Standardization

Several frameworks and methodologies exist to keep track of sustainability performance, but an internationally accepted comprehensive approach is missing [13]. Veleva et al. [14] argue that regarding the problems some have in defining sustainable development, problems in measuring sustainability can be expected to be even greater. Approaches such as GRI seem reasonable within the boundaries of one organization; the application to network performance, though, is not envisaged. Determining the greenhouse gas emission of one organization, for example, does not say anything about the emissions caused by operations upstream the supply chain. For stakeholders, a false impression of the sustainability of the organization may be the result. For instance, customers wanting to buy sustainable products might in fact do the opposite when certain assembly components of the presumably sustainable product are transported between several stations of the globally dispersed manufacturing network or require technological processes prone to high scrap rates [15]. In particular, transportation of goods is left out of sustainability considerations; with regard to the estimation that transportation accounts for 25 % of worldwide CO_2 emissions [16], neglecting transport in sustainability considerations cannot be an acceptable approach. The EFQM has included the full scale of sustainability aspects in the 2010 version of its EFQM Excellence Model ("Taking Responsibility for a Sustainable Future"). The concept of EFQM's self-evaluation, however, is suited best for individual organizations rather than for complex networks and thus provides no advantage over GRI in terms of sustainability performance measurement. This problem of "shifting the burdens" that is inherent to approaches with organizational rather than network focus has been considered in Life Cycle Assessment (LCA) to a certain extent [17]. LCA, however, is a framework the purpose of which is to analyze the status quo within existing business model boundaries, which does not contribute to the reconsideration of basic assumptions and out-of-the-box thinking. Furthermore, LCA does not aim at a full sustainability picture but focuses on environmental impacts while social aspects needed to be considered through complementary tools [18]. In this context, some companies such as the sport lifestyle company "PUMA" have adopted Sustainability Scorecards to keep track of their sustainability performance. Here as well the approach has solely been focused on greenhouse gas emissions and water usage. It seems that social sustainability in general is the category that is least developed in many frameworks [5]; Koho et al. [19], who included political and technological aspects in their 'STEEP' approach, demonstrate that in their approach as well social aspects are less emphasized than economical and environmental [19]. New business models, in the present context, may provide significantly better control over decisive value-adding process steps and life-cycle stages of a product, i.e., new business models have to be evaluated with respect to economical as well as environmental and social aspects. Moreover, in a modern competitive performance context, traditional business tools such as cost-based performance analysis do not provide the necessary space and value creating weight on intangible assets and achievements. They thus may be disqualified for holistic business strategies and performance management applications that include sustainability aspects. The status quo is characterized by strong diversity between different approaches. Existing frameworks differ in their scope (the kind and the number of metrics utilized), their implementation (incorporated in business model or organizational strategy or merely a separate tool), their line of reference (focus on different aspects of sustainability, e.g., focus on environment or focus on economical sustainability), the assessment (self-assessment or assessment through external institutions), their boundaries (organizational focus or life-cycle view), and their purpose (reporting to external stakeholders or information base for management decisions, cf. [20]). Additionally, the importance or the faith that organizations attribute to certain metrics differ [12] which may affect the diligence in measurements.

31.3 Roadmap

31.3.1 Common Sustainability Framework Design and Implementation

Foregoing discussion clarifies the need for a mechanism to cope with modern performance complexities. To achieve high adaptability within complex networks, multi-objective sustainability performance requirements need to be defined and incorporated in a governing framework covering metrics to measure and tools to improve and to verify sustainability and performance.

An important factor for the success of such an approach is adequate guidance and a reference frame for organizations which choose to improve their sustainability performance. Sustainability guidelines will provide orientation, methodologies, techniques and solutions which organizations can employ to improve their results. The objective for sustainable business behavior must be to reduce adverse environmental and social impact while keeping a good economical outcome or vice versa—or improving all three pillars of sustainability at the same time. Hence, Pareto improvement has to be the goal. Dynamic production networks pose a challenge in this context as they involve tradeoffs, e.g., between express deliveries of missing assembly components, environmentally friendly transportation modes and even other implicit business transactions. Those conflicts realistically exist and cannot be ignored. Reinhard [21] criticized early on that the discussion about benefits from sustainable business conduct is polarized with advocates arguing that sustainability pays (e.g., [6] or [19]) and opponents who highlight diminished competitiveness. In cases of tradeoffs, the overall effect among the three pillars of sustainability has to be considered. This can only be reached if all areas of sustainability are incorporated in one governing framework and measured in metrics that can be compared to each other. Such a framework would support organizations' ability to make risk-informed decisions in tradeoff situations and improve commensurability of sustainability KPIs among organizations.

Different authors have examined barriers for the adoption of environmentally friendly measures in manufacturing. Some distinguish between industry barriers and organizational barriers; others call them internal and external barriers. Internal barriers are obstacles which emerge within the organization and adversely affect its ability to implement the measures while external barriers emerge outside the organization with the same consequence [22]. For SMEs in Spain, Murillo-Luna et al. [23] found out that both kinds of barriers play a role for companies; internal barriers such as limited financial capabilities, low employee involvement in decision-making, lack of technological information and communication capabilities, aversion to innovation, and deficient R&D resources, however, are the decisive reasons why companies fail to incorporate sustainability measures. This finding is in accordance with Hillary [22] who conducted an extensive study review and identified internal and external barriers to the implementation of environmental management systems (EMS). While Bunse et al. [24] could determine that industry is lacking support through tools and KPIs when trying to benchmark energy efficiency performance, other authors (e.g., [14]) rather state that numerous indicators already exist and the problem rather lies in missing support how to apply them in order to make the organization more sustainable. In a study comparing performance of 55 companies that adopted sustainability measures and are listed in the Dow Jones Sustainability Index (DJSI) and 55 companies that have not adopted sustainability measures and are not listed in DJSI, López, Garcia, and Rodriguez [25] found out that from 1998 to 2004 there has been no positive correlation between sustainability measures and business performance. On the one hand, as the authors point out, this highlights the problem that some companies may not be willing to adopt sustainability measures when business performance of peers that did so deteriorates rather than benefits. On the other hand, this underlines the need to keep a long-term view on sustainability considerations that is not based around quick payoffs.

31.3.2 Critical Success Factors

To make such a framework successful, certain requirements have to be taken into consideration in different phases of a reference framework.

1. Small and medium-sized enterprises (SMEs; by the EU definition: less than 250 employees and less than 50 million Euro in sales) represent the largest share of enterprises in the European Union and play an important role in manufacturing clusters [26]. Hence, they do play an important role in the effort to create cross-industry sustainable economies. Any approach striving to reach a substantial number of organizations needs to suit SMEs' needs. In particular, this includes the costs for and demands on the organizations when determining sustainability metrics. As Hillary [22] points out, though, there may be decisive differences between micro companies with just a few employees and those with 249 people employed in terms of environmental issues they are facing and capabilities to implement sustainability measures—"and yet they are lumped together in the SME sector".
2. Advantages employing the sustainability frameworks must be clearly visible to the participating organizations. SMEs cannot be expected to employ measures the outcome of which they cannot identify. The framework must incorporate incentives for SMEs for sustainable behavior, such as databases that allow anonymous benchmarking for decisive metrics. Hillary [22] claims that skepticism as to the benefits that can be gained is the main reason for SMEs to not implement measures for environmental improvements. This "culture of inaction" can be an important obstacle for any new sustainability effort.
3. The assessment of sustainability metrics must be as clear as possible and supported by useful guidelines in order to enable maximum commensurability. Drawing on several sources, Lydenberg et al. [27] point out that the number of

companies voluntarily reporting on sustainability performance has reached a significant share in many countries. They do also remark, though, that companies voluntarily reporting often report for different time periods, with different indicators and in different formats, which makes comparison between sustainability performances unfeasible.

4. The framework must allow for different dynamics and interaction levels of different organizations, e.g., globally dispersed production networks as well single suppliers with regional clients.
5. Existing standards and methodologies need to be utilized to the greatest extent possible; well established and working approaches can be used while known weaknesses shall be avoided. Re-inventing existing standards from scratch under a different "brand name" would not contribute to the acceptance of the approach.
6. It must be possible to evaluate new business models and value networks with respect to their environmental, social, and industrial performance which may either be complementary or conflicting.
7. Different KPIs correlate with each other. The vertical and lateral relations have to be understood and utilized in order to derive meaningful conclusions. Systems are made up of several parts that are not isolated but closely interlinked; a random set of elements does not characterize a system [1]. Yet, this linkage between indicators is missing in many frameworks [15]. Hence, a set of KPIs, regardless of its breadth, cannot represent the performance of a system without KPIs being interlinked with each other.

31.3.3 Approach

In order to align the performances of multiple agents in complex production networks for sustainability improvements, a common understanding of sustainability has to be reached early on. Equally important, a common way of measuring sustainability across all different roles and responsibilities has to be introduced to enable individual organizations within those networks to track, to benchmark, and to improve sustainability results. For instance, measuring energy efficiency in the production and delivery of products and services is the basis for controlling energy consumption and for deciding about and tracking of improvements in energy efficiency [24]. Notably, requirements on sustainability measurements differ between globally dispersed production networks and regional production clusters. A standardized approach must therefore allow for different dynamics and different levels of complexity in transactions and supply relations while keeping a value chain focus (Fig. 31.1).

In a first step, core objectives of sustainability performance requirements will be reviewed. This task will be carried out taking into consideration the

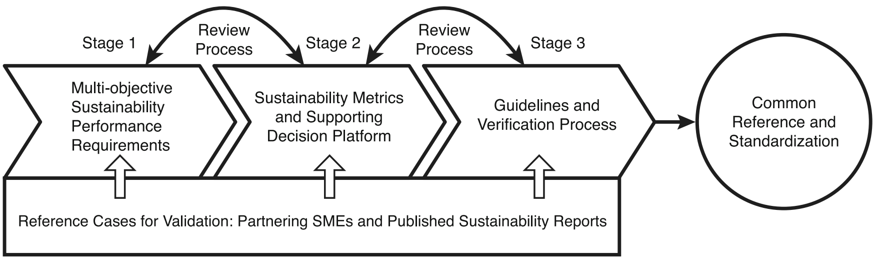


Fig. 31.1 Roadmap for common reference and standardization

complementary and mutually competing factors of sustainability performance in complex production value networks, e.g., possibly evolving tradeoff situations between economical, environmental, and social performance criteria, on order to achieve an overall sustainability improvement. Existing standards and initiatives will be analyzed with respect to their strengths and weaknesses in order to create the knowledge base for the new framework. As Shuaib et al. [5] point out, there may be different perceptions as to the meaning and relevance of sustainability factors. A common understanding, however, is vital to the optimization of production clusters under sustainability considerations [5]. Drawing on the results from this analysis, criteria for sustainability performance measurement will be determined. These will be taken from or created based upon existing metrics from widespread standards and initiatives while expanding the organizational focus to a value network view in order to meet the requirements of complex production clusters. The metrics will be incorporated into the framework. Equally important, guidelines will be defined to support companies in improving their sustainability performance.

31.4 Conclusion

Sustainability has received enormous scientific attention in recent years as indicated by the number of related publications and research projects. A broad variety of frameworks, tools, and standards has evolved in response in order to enable organizations to take sustainability into consideration in business decisions. Sustainability performance has become an important part of corporate reporting for many companies. Nonetheless, a framework equally governing all aspects of sustainability within value network boundaries (as opposed to organizational boundaries) is still missing. *SustainValue* aims at the creation of such a framework by building on the strengths of existing initiatives and taking into account existing hurdles to the implementation of sustainability performance frameworks. The objective is to enable companies to determine their sustainability performance on a standardized basis that allows benchmarking within and between value networks.

References

1. Seiffert MEB, Loch C (2005) Systemic thinking in environmental management: support for sustainable development. J Cleaner Prod 13(12):1197–1202
2. Bretzke WR (2010) Der schwierige Weg zu einer nachhaltigen Logistik. Logistik für Unternehmen (10/2010), p 3
3. SustainValue Consortium (2010) Proposal for a collaborative project: SustainValue
4. European Commission (2009) Work programme 2010—nanoscience, nanotechnologies, materials and new production technologies (NMP). Available from: http://ec.europa.eu/research/participants/portal/page/cooperation?callIdentifier=FP7-NMP-2010-SMALL-4. Cited 11 Aug 2011
5. Shuaib M, Metta H, Lu T, Badurdeen F, Jawahir IS, Goldsby T (2011) Design and performance evaluation of sustainable supply chains: approach and methodologies. In: Seliger G, Khraisheh MM, Jawahir I (eds) Advances in sustainable manufacturing. Springer, Berlin, pp 347–352
6. Porter ME, van der Linde C (1995) Green and competitive: ending the stalemate. Harvard Bus Rev 73(5):120–134 (Sept–Oct)
7. Supplier Business Ltd (2009) Supplying Daimler http://www.supplierbusiness.com/shop/products/21/supplying-the-oems
8. Porter ME (1985) Competitive advantage: creating and sustaining superior performance. Free Press, New York
9. Urban G (2007) Das ganze Zuliefernetzwerk im Griff—Innovations- und Effizienzpotenziale nutzen. Von der Zulieferkette zum Zuliefernetzwerk. In: Logistik in der Automobilindustrie. Innovatives Supply Chain Management für wettbewerbsfähige Zulieferstrukturen; mit 6 Tabellen. Springer, Berlin. pp 1–4
10. Wernerfelt B (1984) A resource-based view of the firm. Strateg Manag J 5(2):171–180
11. Freeman RE (1984) Strategic management: a stakeholder approach. Pitman, Boston
12. Liyanage JP (2003) Operations and maintenance performance in oil and gas production assets: theoretical architecture and capital value theory perspective. Ph.D. Thesis. University of Science and Technology, Norway
13. OECD (2009) Overview of the OECD sustainable manufacturing Toolkit *prototype*. Available from: http://www.oecd.org/dataoecd/34/25/44280332.pdf. Cited 3 Aug 2011
14. Veleva V, Hart M, Greiner T, Crumbley C (2001) Indicators for sustainable production. J Cleaner Prod 9(5):447–452
15. Rabe M, Jochem R, Weinaug H (2011) Multi-perspective modelling of sustainability aspects within the industrial environment and their implication on the simulation technique. In: Seliger G, Khraisheh MM, Jawahir I (eds) Advances in sustainable manufacturing. Springer, Berlin, pp 315–320
16. International Energy Agency (2009) Transport, energy and CO2. Moving toward sustainability. OECD Publishing, Paris
17. European Commission (2010) International reference life cycle data system (ILCD) handbook: general guide for life cycle assessment—detailed guidance. 1st edn. Publications Office of the European Union, Luxembourg
18. Trusty W (2010) Misconceptions and misunderstanding about LCA. Athena Sustainable Materials Institute, Merrickville
19. Koho M, Nylund H, Arha T, Torvinen S (2011) Towards manufacturing system sustainability assessment: an initial tool and development plans. In: Seliger G, Khraisheh MM, Jawahir I (eds) Advances in sustainable manufacturing. Springer, Berlin, pp 309–314
20. Feng S, Joung C (2010) An overview of a proposed measurement infrastructure for sustainable manufacturing. In: Proceedings of the 7th global conference on sustainable manufacturing. Chennai, 2–4 Dec 2009
21. Reinhard F (1999) Bringing the environment down to Earth. Harvard Bus Rev 77(4):149–157

22. Hillary R (2004) Environmental management systems and the smaller enterprise. J Cleaner Prod 12(6):561–569
23. Murillo-Luna JL, Garcés-Ayerbe C, Rivera-Torres P (2011) Barriers to the adoption of proactive environmental strategies. J Cleaner Prod 19(13):1417–1425
24. Bunse K, Vodicka M, Schönsleben P, Brülhart M, Ernst FO (2011) Integrating energy efficiency performance in production management—gap analysis between industrial needs and scientific literature. J Cleaner Prod 19(6–7):667–679
25. López MV, Garcia A, Rodriguez L (2007) Sustainable development and corporate performance: a study based on the Dow Jones sustainability index. J Bus Ethics 75(3):285–300
26. European Commission (2011) Key figures in European business with a special feature on SMEs http://bookshop.europa.eu/en/key-figures-on-european-business-with-a-special-feature-on-smes-pbKSET11001/
27. Lydenberg S, Rogers J, Wood D (2010) From transparency to performance: a new method for industry-based sustainability reporting on key issues. Initiative for Responsible Investment at Harvard University. Available from: http://hausercenter.org/iri/wp-content/uploads/2010/05/IRI_Transparency-to-Performance.pdf. Cited 23 Aug 2011

Chapter 32
Modeling the Impact of Working Capital Management on the Profitability in Industrial Maintenance Business

S. Marttonen, S. Viskari and T. Kärri

Abstract In this paper, we present an analytical model for flexible asset management, which is a new tool for company decision-making. The model reveals a significant negative correlation between the cycle times of operational working capital and the return on investment. Conventional research on working capital management has mostly focused on manufacturing industries. We show that working capital should be managed actively also in unconventional environments like service industries. The focus is on the industrial maintenance service providers, which still remain somewhat unexplored in academic literature. The importance of working capital management is actually emphasized in this industry, due to its light fixed assets and good profitability. Interestingly, there are some major differences between large enterprises and small and medium size enterprises in the industrial maintenance service sector. These can be explained through economies of scale and the fact that large maintenance service enterprises often focus on providing services mostly for their former host companies.

32.1 Introduction

The focus of this paper is on the working capital management of industrial maintenance service providers. There exists little previous research about industrial maintenance service providers, although maintenance performance has grown

---

S. Marttonen (✉) · S. Viskari · T. Kärri
Industrial Management, Lappeenranta University of Technology,
53851 Lappeenranta, Finland
e-mail: salla.marttonen@lut.fi

S. Viskari
e-mail: sari.viskari@lut.fi

T. Kärri
e-mail: timo.karri@lut.fi

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_32,

to be of high importance to industrial companies. Previous research concerning industrial maintenance services has been mainly done from the point of view of service buyers, concerning mostly outsourcing. The present state of industrial services, the global trend toward networking and the focus on core competences has led to some notable changes. A vast majority of outsourced business functions can be classified as services [1–3]. Bailey et al. [1] reported that 70 % of the companies in their research sample had outsourced one or more of their business functions. Gradually the outsourcing practice has extended to more and more important functions [4]. Consequently an increasing demand for a variety of industrial services has emerged.

Because of the important role of maintenance performance to the industrial companies, long-term partnerships are often preferred with the maintenance service providers. This highlights the role of contracts in maintenance, which have been addressed in scientific research e.g., [8–11]. However, the perspective of the maintenance service provider remains somewhat unexplored, and the perspective of the customer has been strongly emphasized instead. The focus has mainly been on determining the operational responsibilities of the parties and on specifying rewards and sanctions for the service provider. The maintenance service providers must have in advance a good idea of how changes in the contract affect their profitability. In addition to the responsibilities, rewards and sanctions mentioned above, elements to be considered include for example the terms of payment and the ownership of spare parts and assets to be maintained. These can be further inspected with the flexible asset management model introduced in this paper.

The role of maintenance is significant and still increasing in Finland due to relatively aged industrial assets: a vast majority of the productional investments of Finnish industrial companies are nowadays directed to countries with inexpensive work force [12]. Industrial maintenance outsourcing is relatively extensive in Finland. The Finnish Maintenance Society [13] states that out of the 50,000 people employed in industrial maintenance in Finland, 15,000 are employed by maintenance service providers. According to this indicator, 30 % of Finnish industrial maintenance has been outsourced. The recent profitability problems of the forest industry have further increased maintenance outsourcing in Finland. Forest companies have struggled to cut down the costs and to outsource anything but their most crucial core functions.

Working capital is a part of a firm's capital employed, such as fixed assets. Broadly, net working capital is defined as current assets less current liabilities, and it measures the financial health of a company. Working capital can also be defined from the operational perspective as inventories plus accounts receivable less accounts payable. We concentrate on operational working capital in this paper. Although discussed in textbooks of finance and management accounting, working capital management has received relatively little attention in the academic literature [14]. The research has traditionally concentrated on long-term capital [15], and asset management has mostly concerned fixed assets [16]. The past economic crisis raised the interest toward more efficient working capital management. This also increased the academic interest toward working capital management.

Operational working capital research has basically discussed two issues: the determinants of working capital management and the relation between working capital management and profitability have been studied by several researchers e.g., [15, 17, 18]. Statistical analysis of the relation between working capital management and profitability have concluded that it is more beneficial for a company to aim for a shorter cycle time of operational working capital and reduce tied-up capital and total assets than to operate with long cycle times e.g., [18–20].

Previous studies have used statistical analysis tools to examine the relationship between the cycle time of working capital and relative profitability. The data of the studies has been gathered from large databases, which has enabled large samples. With this kind of research design, statistically significant results can be found, but the previous literature lacks the perspective of individual industries. As we know, there are huge differences in the capital structures, investment logics and profitability between industries. In addition, it is still unknown how great the impact of working capital management on profitability actually is. There is still not a clear view of what the variables affecting working capital are and how great their impact is. According to Reed and Storrud-Barnes [21], most management theories have concentrated on manufacturing companies, while differences and similarities to service providers have been neglected. Thus the applicability of practical contributions to the service sector often remains unsettled. In many respects this holds true also in studies of working capital management. In some studies of working capital management, service providers have been excluded from the sample in advance see e.g., [18, 22]. Other studies have included service providers alongside manufacturing companies see e.g., [23, 24]. It can be noticed in these studies that the cycle time of operational working capital, cash conversion cycle (CCC), appears to be significantly shorter for service providers than for manufacturing industries. However, this observation has not yet been discussed further, and thus it is reasonable to conduct further research on this subject. Both in consumer and industrial markets, the role of services has grown to be massive. In this paper, modeling has been used as the research method. The behavior and features of the model are tested with empirical data of industrial maintenance service providers. The objective is to examine the role of working capital and its impact on the return on investment (ROI) in the industrial maintenance service sector. The research questions of this paper are the following:

- *What is the effect of working capital management on the return on investment in the industrial maintenance service sector?*
- *What are the main differences between SMEs and large enterprises in the maintenance service sector regarding the cycle times of working capital?*

32.2 Research Design

Analytical modeling has been used as a research method in this study. Demski [25] describes this method as using deductive logic in representing a concept or a process. The advantage of using modeling is transparency, which leads to high internal validity. Kärri [26] has constructed a model for forecasting the investment needs (consisting of fixed assets and net working capital) of a company or an industry under changing sales. Kärri argues that companies have or should have a long-term target ROI-level. When the fluctuation of sales is great, the capital structure should be flexible. This means that investments on the fixed assets but also investments on the working capital should be tied down to sales. Kärri's model observes the net working capital rate, but the flexible asset management model used in this paper (henceforth called the FAM model) highlights working capital management when cycle times are added to the model. The derivation of the FAM model is presented in Appendix A. The two forms of the FAM model used in this paper are presented in Eq. (32.1). The ROI is determined by five parameters. Besides the cycle time of operational working capital (CCC), also the profit margin (EBITDA %), the fixed asset ratio (FA %), the average depreciation period (B) and the cycle time of residual (r), consisting of other current assets less other current liabilities, affect the relative profitability. The CCC can be further divided into three components: the cycle time of inventories (DIO), the cycle time of accounts receivable (DSO), and the cycle time of accounts payable (DPO). The ROI increases if *ceteris paribus* the EBITDA % or the B increases, or the FA %, the CCC or the r decreases. Equation (32.1) works correctly if both the numerator and the denominator are positive and the B has a value greater than 1. It should be noted that asset management is a broad concept which is not limited to working capital management. This paper focuses on working capital, but the FAM model can be applied to many different asset managerial situations in future research.

$$\mathrm{ROI} = \frac{\mathrm{EBITDA}\,\% - \left(\mathrm{FA}\,\% \cdot \frac{1}{\mathrm{B}-1}\right)}{\frac{\mathrm{CCC}}{365} + \frac{\mathrm{r}}{365} + \mathrm{FA}\,\%} = \frac{\mathrm{EBITDA}\,\% - \left(\mathrm{FA}\,\% \cdot \frac{1}{\mathrm{B}-1}\right)}{\frac{\mathrm{DIO}}{365} + \frac{\mathrm{DSO}}{365} + \frac{\mathrm{DPO}}{365} + \frac{\mathrm{r}}{365} + \mathrm{FA}\,\%}. \tag{32.1}$$

The FAM model is applied to industrial maintenance business by using an empirical analysis about the behavior of the model. The more extensive testing of the model with different data is left for further research. Though small, the analyzed sample of 18 service providers represents a remarkable share of the Finnish industrial maintenance sector. The 3,893 people employed by these 18 enterprises cover approximately 26 % of the employees of the whole sector. The sample net sales of 472 million euro, on the other hand, represent approximately 13 % of the sum used yearly in industrial maintenance in Finland. The enterprises selected to the sample are listed in Appendix B, together with their personnel numbers and net sales. The sample has been selected by utilizing the membership list of the Finnish

Maintenance Society. Structurally, the sample represents the Finnish industrial maintenance sector well, consisting of a few large enterprises and mostly SMEs. The sample has been divided into large enterprises and SMEs according to the European Commission recommendation 2003/361/EC [27]. Maintenance companies from outside Finland were excluded from the study because the required data was not within reach. Probability sampling has not been used, because a complete list of suitable enterprises was not available. It was considered crucial that the sample especially reflects the features of industrial maintenance business. Thus, when selecting the sample, only enterprises whose focus was mainly in industrial maintenance were accepted, which means that equipment manufacturers and enterprises offering infrastructural maintenance services were delimited from the sample. Also the smallest micro enterprises were left out of this research. The final criteria of belonging to the sample were that the enterprises had existed for the whole period of analysis, 2004–2008, and their financial statements were available for this period. At the time of collecting the data, the financial statements of subsequent fiscal years were not yet available through the chosen database. The financial statement data was collected using the Voitto + database (ISSN 1459-9457). This database is maintained by Suomen Asiakastieto Ltd. (loose translation Finnish Customer Information), which is a leading service provider in Finland producing corporate information, which is broadly considered to be reliable and objective. Given the sampling criteria discussed above, five large enterprises suitable for the analysis were found. Considering that the maintenance sector still evolves quite rapidly, a more extensive sample of large maintenance providers can be selected probably in a few years' time. 13 maintenance SMEs were included in the analysis in order to reach a representative sample. Large enterprises and SMEs are mostly analyzed separately, but the average values of the whole sample are also used. It is justified to have more SMEs than large enterprises in the sample to simulate the actual structure of the maintenance industry.

32.3 The Importance of Working Capital Management in the Maintenance Service Sector

The impact of working capital management on profitability is presented in Figs. 32.1 and 32.2 by changing the CCC and setting the other parameters of Eq. (32.1) as constant in the FAM model. The CCC gets values from zero to 100. We have examined how the length of the CCC affects the ROI on different levels of the FA %, the EBITDA %, the B and the r. The constant values are industry averages in the industrial maintenance service sector calculated with the data used in the empirical analysis. The range of each variable in Figs. 32.1 and 32.2 has been found realistic in the observed industry.

If the fixed asset ratio is high, the effect of the cycle time of working capital on profitability remains small. The impact of the CCC on profitability is bigger in

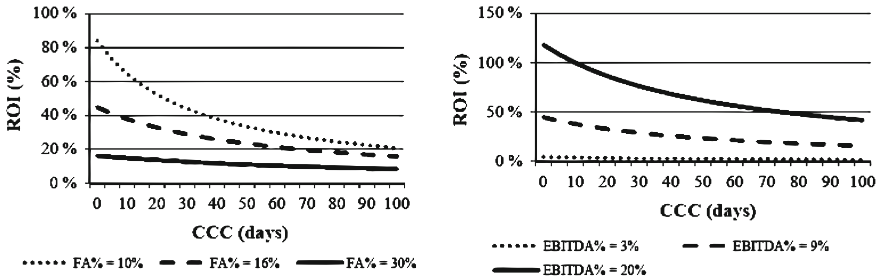


Fig. 32.1 The ROI as a function of the CCC and the FA % (*left*) and as a function of the CCC and the EBITDA % (*right*). Here the B equals 8 years and the r equals 4 days

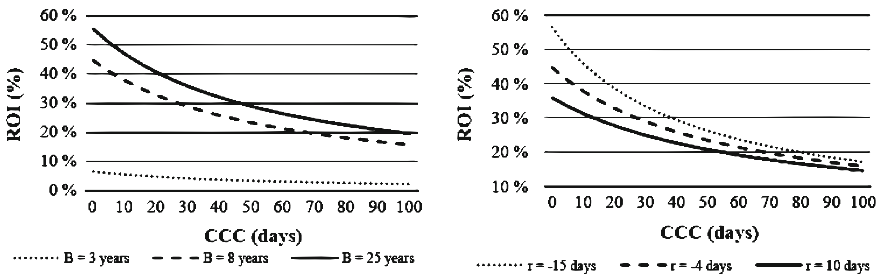


Fig. 32.2 The ROI as a function of the CCC and the B (*left*) and as a function of the CCC and the r (*right*). Here the EBITDA % equals 9 % and the FA % equals 16 %

companies with flexible asset structures and a lower fixed asset ratio, and thus they should pay more attention to working capital management. For comparison, in a big customer industry of the industrial maintenance service sector, the pulp and paper industry, the FA % is typically over 100 %. The length of the CCC gets more important when the EBITDA % grows. If the EBITDA % is low, efficient working capital management cannot save profitability, but on higher levels of the EBITDA, a long CCC can destroy profitability. Since a low fixed asset ratio increases the impact of the CCC on profitability, it is only logical that the same thing happens when the average depreciation period shortens. On the other hand, we can see that the shorter the residual term is, the more the CCC affects the ROI. As regards the CCC, the FA %, the B and the r, it must be noted, however, that minimizing them is not always the most viable solution. If anything, these parameters should be optimized on the basis of operational, tactical and strategic criteria. Especially the residual term, consisting of other current assets and other current liabilities, calls for further scientific research before any specific comment about the optimal value can be made. Based on this analysis we can now come to a conclusion that the characteristics of the industrial maintenance service sector strongly support the importance of working capital management practices. Operating with light fixed assets and high profitability makes these enterprises vulnerable to changes in profitability caused by fluctuation in the CCC.

32.4 Comparing Large Enterprises and SMEs in Industrial Maintenance Service Sector

The parameters for the FAM model for all the enterprises in this research, large enterprises and SMEs separately, have been calculated on the basis of the financial statements of the companies. These parameters are presented in Table 32.1. It can be noticed that the EBITDA % is significantly higher in SMEs than in large maintenance service providers. However, when profitability is proportioned to the capital invested and the ROI is inspected, the situation changes. This indicates that the large maintenance service enterprises operate with exceptionally light balance sheets. For SMEs, the CCC is approximately 53 days, whereas for large companies the corresponding figure is roughly 9 days. When reviewing the components of the CCC in Table 32.1, it can be concluded that this difference mainly originates from the SMEs having notably longer cycle times for both accounts receivable and inventories. This may signal the existence of considerable economies of scale in the industrial maintenance sector. Large companies may have been able to shorten their DSO by using their bargaining power over their customers, whereas the bargaining power of SMEs over their customers is minor. Considering the DIO, on the other hand, the economies of scale in the industrial maintenance sector are quite different. Generally speaking, large maintenance service providers have more customers, facilities and equipment to be maintained, compared to SMEs. The risk of asset breakdown is thus usually divided between several units. It should also be noted that the FA % is significantly larger for SMEs than for large enterprises. This may be due to economies of scale related to the use of fixed assets. Previous research suggests that economies of scale truly exist in the industrial maintenance business [10, 28, 29].

Besides economies of scale, there is another explanation for the differences in the parameters between large enterprises and SMEs. It is notable that large enterprises in the Finnish industrial maintenance service sector are mostly focused on customers operating in specific industries. In fact many of the large service providers have been founded by manufacturing enterprises, who have wanted to outsource their maintenance. This has led to a situation where a vast majority of the net sales of most large maintenance service providers comes from their host companies. For SMEs, this kind of situation is not so common. Complete conclusions are difficult to make from this basis, but close cooperation with their host companies may have had an influence on the studied parameters of large maintenance enterprises. At least the cycle time of accounts receivable, the cycle time of inventories, the fixed asset ratio and the residual depend on the contractual relation between host companies and service providers. The host companies may agree to short payment times, which would shorten the DSO. Some part of the inventories and the fixed assets of maintenance service providers can be owned and stored by their host companies, which reduces the DIO and the FA %. Again, a large amount of receivables from group companies increase the sum of other current assets, which finally makes the r larger. It would seem that the operations

Table 32.1 Parameters of industrial maintenance service providers for the FAM model (average numbers from 2004 to 2008)

	EBITDA (%)	FA (%)	B (years)	DSO (days)	DIO (years)	DPO (years)	CCC (days)	R (days)	ROI (%)
Large enterprises	5	8	6.9	15	9	15	9	7	30
SMEs	10	20	7.8	45	22	14	53	−10	23
All enterprises	9	16	7.5	37	18	14	41	−6	24

and the financial statements of the large maintenance service providers are affected by both the economies of scale and their host companies. It would be very important to know how strong impacts each of these aspects have, but this is left for further research.

The EBITDA %, the FA %, the B and the r have now been set constant in the FAM model for large maintenance service providers and SMEs separately. The studied variables have thus been limited to the CCC and the ROI. We acknowledge the great impact of the other four parameters on the ROI, but analyzing this impact exceeds the scope of this paper. We have varied the length of the CCC in order to define the impact on the ROI. Figure 32.3 shows how the ROI changes for SMEs and for large enterprises. The actual situation for each group is depicted with the vertical lines. The difference between SMEs and large enterprises is remarkable considering the length of the CCC. Arising from the shape of the relation between the CCC and the ROI, this means that for large enterprises changes in the CCC have a more extensive effect on the ROI. Thus particularly the large industrial maintenance service providers should pay attention to working capital management. Also Hatinen et al. [30] remark that working capital management is more important for large maintenance service providers than for SMEs. They have come to this conclusion by noticing that in large maintenance enterprises the share of working capital investments in total investments is remarkably larger than in SMEs.

Figure 32.4 presents the development of the CCC and the ROI during the period of analysis for large enterprises and SMEs. The values for the ROI in Fig. 32.4 have been calculated with the FAM model. The EBITDA %, the FA %, the B and the r have been set constant for each year and each group of enterprises separately. These constant values equal the yearly average numbers for each group of enterprises. It can be seen that in large enterprises the CCC has clearly become longer, while the ROI has decreased. The CCC is still exceptionally short, but the trend has been somewhat alarming. It is crucial to stop this trend in order to keep the good profitability of these enterprises. The focus should be on the terms of

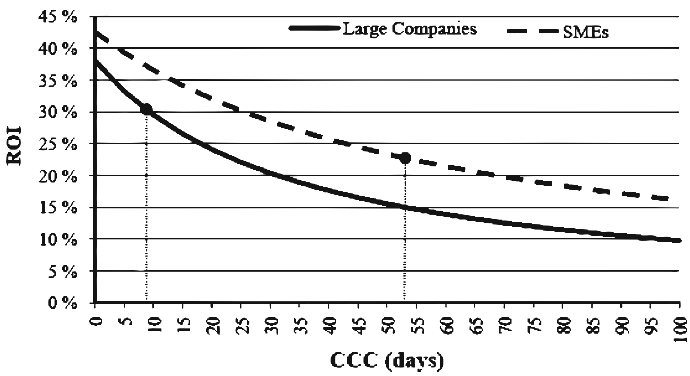


Fig. 32.3 The impact of the CCC on the profitability of large companies and SMEs in the maintenance industry. The *vertical lines* indicate the actual situation of the enterprises

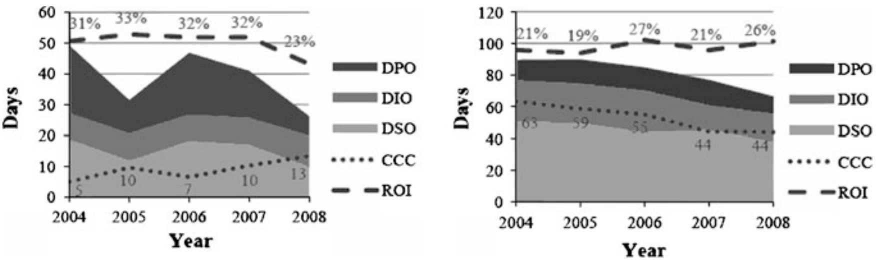


Fig. 32.4 Trends of the CCC and ROI development in large maintenance enterprises (*left*) and in maintenance SMEs (*right*)

payment. Both the DSO and the DPO of the large enterprises have shortened. However, the decrease in the DPO significantly outdoes the positive trend of the DSO. After reviewing the figures behind the DPO, it can be concluded that the accounts payable of the large enterprises have fluctuated violently during the period of analysis. On the other hand, both the purchase during the fiscal year and the inventories have developed more steadily. No definite conclusions can be drawn from public financial statements, but it would seem that the purchase of industrial maintenance service providers spreads very unevenly throughout the fiscal year. This could mainly result from the characteristics of the industrial maintenance business. The safety stock of spare parts is of utmost importance, and small-sized orders to suppliers are placed at irregular intervals. In SMEs the CCC has shortened and the ROI has increased during the period of analysis, so the trend is positive. Both the DSO and the DIO have shortened significantly for SMEs. The development has been steadier for SMEs than for large enterprises. The DSO may have shortened because of longterm relations to customers. However, it should be noticed that also the characteristics of the industrial maintenance business affect the uneven accumulation of the accounts receivable: even though the role of preventive maintenance has increased substantially, reactive maintenance actions still give a lot of work to industrial maintenance enterprises. The shortening of the DIO can be partly explained through the economies of scale. As the enterprises have grown year by year, the risks have been divided between more and more units. Thus the ratio of spare part safety stock to net sales has supposedly decreased. In addition, the value of the inventories in SMEs has fluctuated a lot, which has affected the development of the DIO. This may reflect the fact that the inventory management of SMEs is often not very organized.

32.5 Conclusion

The managerial implications of our FAM model are extensive. The model can be utilized as a tool in decision-making in firms, both in the short term and in the long term. In contract-bound industries the FAM model can be used when negotiating with customers or suppliers to point out any changes in contract profitability caused by for example different terms of payment or arrangements of the ownership of assets. In the long term, decision-makers can use the model to discover any positive or negative trends in the development of their firm, and finally to forecast the future development. On the basis of this paper, the decision-makers of companies can consider how important working capital management is in their industry. In addition, we now know that in the industrial maintenance service business attention should be paid to active management of working capital. We can conclude that this holds true especially in large industrial maintenance service enterprises. Previous studies of industrial maintenance companies have not addressed working capital management, which gains more and more attention under the volatile economic circumstances of the present day. There is a growing need for management tools that focus on the impact of working capital management on profitability. The past economic crisis has reshaped the profitability of the maintenance service providers in an unknown way. The FAM model is a valuable management tool especially under this kind of hard circumstances. In order to restore their profitability, companies can either manage their working capital or adjust their fixed assets. It should be noted that while this paper focuses on Finnish maintenance providers, the model and the logic presented here can be applied to a variety of industries and countries.

Our FAM model can be used as a tool for decision-making in any company. However, for the time being, there are some theoretical limitations in the applicability of the model. Solving this problem and applying the model to different asset management situations are natural objectives for further research. The main contribution of this study is analytical, not statistical. The sample used in empirical analysis represents the Finnish industrial maintenance service sector well. However, the scarcity of large enterprises in the market in question must be included in our research limitations. As our analysis of large maintenance service providers was done with only five enterprises, any single abnormalities in numerical values affect the average values excessively. Despite this, the empirical analysis conducted is valid for testing the FAM model in practice. Concerning the generality of our results, it can be said that in addition to the industrial maintenance service business, many of our conclusions also hold true in other industries. For example, the impact of the CCC on the ROI extends while the fixed assets, the average depreciation period or the residual decrease, or when the operating margin ratio increases. This supports the importance of working capital management in unconventional industries, for example in many service industries. On the other hand, the results of comparing the development of large enterprises and SMEs in the industrial maintenance service sector include some contextual elements which somewhat limit the generalization of the results. However, there are also some global elements, such as economies of scale or the general cost and asset structure of industrial maintenance enterprises.

Objectives for further research include discussing the residual term consisting of the other current assets less the other current liabilities. The behavior of this variable should be defined so that decision-makers can optimize its value. Further research can also include international comparative research about applying the FAM model into practice. It should also be further examined how important a role economies of scale truly have in the industrial maintenance business. On the other hand, providing services mostly for their host companies has an impact on the operations and financial statements of industrial maintenance service enterprises. The role of this impact would also be important for both academics and the decision-makers of firms to know.

Appendix A.1 (1/2)

Deriving the Analytical Flexible Asset Management Model

The starting point of the FAM model is the basic function of the return on investment. As financial statements are used as a data source, the model is also derived from the balance sheet items. The ROI is determined as a function of the operating income (EBIT) and the capital invested (C).

$$\mathrm{ROI} = \frac{\mathrm{EBIT}}{\mathrm{C}}. \tag{32.2}$$

The depreciations and amortizations (D) can be separated from the EBIT. If the capital employed is described with the assets of the balance sheet, it consists of the fixed assets (FA) plus the net working capital (NWC), which equals the equity plus long-term loans.

$$\mathrm{ROI} = \frac{\mathrm{EBITDA} - \mathrm{D}}{\mathrm{NWC} + \mathrm{FA}}. \tag{32.3}$$

The net working capital can be divided further to current assets (CA) less current liabilities (CL).

$$\mathrm{ROI} = \frac{\mathrm{EBITDA} - \mathrm{D}}{\mathrm{CA} - \mathrm{CL} + \mathrm{FA}}. \tag{32.4}$$

Further, operational working capital components, inventories (INV), the accounts receivable (AR) and the accounts payable (AP), can be separated from the current assets and the current liabilities, leaving the other current assets (OCA) and the other current liabilities (OCL) consisting of non-operational working capital items, such as the cash and cash equivalents and the current portion of long-term liabilities.

$$\mathrm{ROI} = \frac{\mathrm{EBITDA} - \mathrm{D}}{\mathrm{INV} + \mathrm{AR} - \mathrm{AP} + \mathrm{OCA} - \mathrm{OCL} + \mathrm{FA}}. \tag{32.5}$$

With the help of the fixed assets (FA) and the yearly depreciations and amortizations (D), the average depreciation period (B) can be calculated as follows

$$\mathrm{B} = \frac{(\mathrm{FA} + \mathrm{D})}{\mathrm{D}} = \frac{\mathrm{FA}}{\mathrm{D}} + 1, \tag{32.6}$$

and when turned around

$$\mathrm{D} = \frac{\mathrm{FA}}{\mathrm{B} - 1}. \tag{32.7}$$

Applying this definition of the depreciations and amortizations (D) to the original model gives

$$\mathrm{ROI} = \frac{\mathrm{EBITDA} - \frac{\mathrm{FA}}{\mathrm{B}-1}}{\mathrm{INV} + \mathrm{AR} - \mathrm{AP} + \mathrm{OCA} - \mathrm{OCL} + \mathrm{FA}}. \tag{32.8}$$

Cycle times have been used in many studies to measure the operational working capital management. Many studies use the cash conversion cycle (CCC) to measure the working capital management. We also apply this approach and observe working capital management through cycle times. For this we have to deduce our FAM model further. The CCC is defined through three components, the cycle time of inventories (DIO), the cycle time of accounts receivable (DSO), and the cycle time of accounts payable (DPO)

$$\mathrm{CCC} = \mathrm{DIO} + \mathrm{DSO} - \mathrm{DPO}. \tag{32.9}$$

The cycle times of inventories (days inventories outstanding) can be calculated as the inventories divided by sales (S).

$$\mathrm{DIO} = \frac{\mathrm{INV}}{\mathrm{S}} \cdot 365. \tag{32.10}$$

Similarly, the cycle time of accounts receivable (days sales outstanding) is

Appendix A.2 (2/2)

Deriving the Analytical Flexible Asset Management Model

$$\mathrm{DSO} = \frac{\mathrm{AR}}{\mathrm{S}} \cdot 365, \tag{32.11}$$

and the cycle time of accounts payable (days payables outstanding)

$$\mathrm{DPO} = \frac{\mathrm{AP}}{\mathrm{S}} \cdot 365. \tag{32.12}$$

We can also determine the cycle time of other net working capital components in a similar manner. When the cycle time of other current assets and the cycle time of other current liabilities are combined, we get the residual (r).

$$\mathrm{r} = \frac{\mathrm{OCA}}{\mathrm{S}} \cdot 365 - \frac{\mathrm{OCL}}{\mathrm{S}} \cdot 365. \tag{32.13}$$

In order to calculate the ROI using the cycle time defined above, Eq. (32.8) has to be divided by sales (S). When doing so, we can write the equation using only percentages (EBITDA/S = EBITDA % and FA/S = FA %), and cycle times as parameters

$$\mathrm{ROI} = \frac{\mathrm{EBITDA}\,\% \; - \left(\mathrm{FA}\,\% \; \cdot \frac{1}{\mathrm{B}-1}\right)}{\frac{\mathrm{DIO}}{365} + \frac{\mathrm{DSO}}{365} - \frac{\mathrm{DPO}}{365} + \frac{\mathrm{r}}{365} + \mathrm{FA}\,\%}, \tag{32.14}$$

and when combining the components of the operating working capital, we can write

$$\mathrm{ROI} = \frac{\mathrm{EBITDA}\,\% \; - \left(\mathrm{FA}\,\% \; \cdot \frac{1}{\mathrm{B}-1}\right)}{\frac{\mathrm{CCC}}{365} + \frac{\mathrm{r}}{365} + \mathrm{FA}\,\%}. \tag{32.15}$$

Appendix B

The Enterprises Selected for the Research

(continued)

	Personnel in 2007	Net sales in 2007, M€
Personnel in 2007		Net sales in 2007, M€
Large enterprises		
Fortek Ltd.	858	118.2
Varenso Ltd.	473	83.4
Konecranes Service Ltd.	516	78.9
Botnia Mill Service Ltd.	591	73.9
Kymenso Ltd.	743	45.0
Large enterprises in total	3,181	399.4
SMEs		
ISS Teollisuuspalvelut Ltd.	104	12.5
Tespal Ltd.	53	9.4
Machinery Service Finland Ltd.	92	8.4
Mainpartner Industrial Services Ltd.	109	8.0
Pikoteknik Ltd.	50	7.3
Tormets Ltd.	103	6.3
Mahro Ltd.	28	5.5
Betamet Service Ltd.	48	3.9
Astepa Ltd.	46	3.1
Metso Mill Service Kauttua Ltd.	36	2.6
Kangasalan Pajaservice Ltd.	16	2.6
JTT Konepaja Ltd.	20	2.2
Rauman Sähkökonehuolto Ltd.	10	0.8
SMEs in total	712	78.6
All enterprises in total	3,893	472.0

References

1. Bailey W, Masson R, Raeside R (2002) Outsourcing in Edinburgh and the Lothians. Eur J Purchasing Supply Manage 8(2):83–95
2. Benson J, Ieronimo N (1996) Outsourcing decisions: evidence from Australia-based enterprises. Int Labour Rev 135(1):59–73
3. Kakabadse A, Kakabadse N (2002) Trends in outsourcing: contrasting USA and Europe. Eur Manage J 2(20):189–198
4. Hui E, Tsang A (2006) The inter-organizational relationship in a multi-contractor business network. J Qual Maintenance Eng 12(3):252–266
5. Hilletofth P, Hilmola OP (2010) Role of logistics outsourcing on supply chain strategy and management: survey findings from Northern Europe. Strateg Outsourcing 3(1):46–61
6. Tarakci H, Tang K, Teyarachakul S (2009) Learning effects on maintenance outsourcing. Eur J Oper Res 192(1):138–150
7. The European Federation of National Maintenance Societies (2011) Intermediate results of the asset management survey for all national maintenance societies. Available from: http://e-lomake.fi/raportti/efnms_2011_survey. Cited 25 July 2011

8. Campbell J (1995) Outsourcing in maintenance management: a valid alternative to self-provision. J Qual Maintenance Eng 1(3):18–24
9. Kumar R, Markeset T, Kumar U (2004) Maintenance of machinery: negotiating service contracts in business-to-business marketing. Int J Serv Ind Manage 15(4):400–413
10. Martin HH (1997) Contracting out maintenance and a plan for future research. J Qual Maintenance Eng 3(2):81–90
11. Ngwenyama O, Bryson N (1999) Making the information systems outsourcing decision: a transaction cost approach to analyzing outsourcing decision problems. Eur J Oper Res 115(2):351–367
12. Komonen K (2009) Challenges 2009: the current and future challenges of maintenance in Finland. Promaint 3(23):18–22 (in Finnish)
13. Finnish Maintenance Society (2007) Maintenance in national economy of Finland. Available from: http://www.promaint.net/alltypes.asp?menu_id=348 (in Finnish). Cited 5 May 2011
14. Viskari S, Lukkari E, Kärri T (2011) State of working capital management research: bibliometric study. Middle East Finance Econ 14:99–108
15. Chiou J, Cheng L, Wu H (2006) The determinants of working capital. J Am Acad Bus 1(10):149–155
16. Komonen K (2010) Asset management in the industrial sector: background and conceptual approach. Maintworld 2(1):16–19
17. Hill MD, Kelly GW, Highfield MJ (2010) Net operating working capital behavior: a first look. Finance Manage 39(2):783–805
18. Deloof M (2003) Does working capital management affect profitability of Belgium firms? J Bus Finance Account 30(3–4):573–587
19. Lazaridis I, Tryfonidis D (2006) Relationship between working capital management and profitability of listed companies in the Athens stock exchange. J Financ Manage Anal 19(1):26–35
20. García-Teruel PJ, Martínez-Solano P (2007) Effects of working capital management on SME profitability. Int J Manag Finance 2(3):164–177
21. Reed R, Storrud-Barnes S (2009) Systematic performance differences across the manufacturing-service continuum. Serv Bus 4(3):319–339
22. Dong HP, Su J (2010) The relationship between working capital management and profitability: a Vietnam case. Int Res J Finance Econ 5(49):59–67
23. Baños-Caballero S, García-Teruel P, Martínez-Solano P (2010) Working capital management in SMEs. Account Finance 3(50):511–527
24. Jose M, Lancaster C, Stevens J (1996) Corporate returns and cash conversion cycles. J Econ Finance 1(20):33–46
25. Demski JS (2007) Analytic modeling in management accounting research. In: Chapman CS, Hopwood AG, Shields MD (eds) Handbook of management accounting research. Elsevier Ltd, Amsterdam, pp 365–371
26. Kärri T (2007) Timing of capacity change: models for capital intensive industry. Ph.D. Thesis, Lappeenranta University of Technology, Lappeenranta
27. European Commission (2009) Definition of micro, small and medium-sized enterprises adopted by the commission. Available from: http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:124:0036:0041:en:PDF. Cited 6 May 2011
28. Hecker A, Kretschmer T (2010) Outsourcing decisions: the effect of scale economies and market structure. Strateg Organ 8(2):155–175
29. Komonen K (2002) A cost model of industrial maintenance for profitability analysis and benchmarking. Int J Prod Econ 79(1):15–31
30. Hatinen L, Pirttilä M, Viskari S, Kärri T (2011) Investment logics of Finnish industrial maintenance service providers. Int J Strateg Eng Asset Manage 1(1):33–48

Chapter 33
From-Design-to-Operations Risk Mitigation in Nordic Wind Energy Assets: A Systematic RAMS+I Management Model

R. Tiusanen, J. Jännes and J. P. Liyanage

Abstract The offshore wind energy sector is growing fast and a lot of effort is put to the business and technology development. Today most current offshore turbines are based on onshore turbine designs and the turbine technology development continues incrementally. At the same time offshore wind marked demands for technological breakthroughs and new concepts increases. The main focus in offshore wind turbine design is reliability and cost efficiency. Most important decisions concerning safety and dependability are made in the early phases of design process that has implications in the operational phase. The later the decisions are made and mistakes realized the more they cost. RAMS+I (Reliability, Availability, Maintainability, Safety and Inspectability) issues should be taken into account systematically from the beginning of design process as a risk mitigating measure. Wind turbines in cold climates like in Nordic countries may be exposed to icing conditions or temperatures outside the design limits of standard wind turbines. In this paper we outline a model for managing RAMS+I factors in the conceptual design phase of offshore wind turbine. The model is based on the product development process, concurrent design ideas and the Stage-Gate® model. The model concentrates mostly on technical decisions made in the early development phases. Service aspects, such as maintenance concepts, are not in the focus of this paper. In some level service-based factors still have to be involved into technical decisions made. The objective of the presented model is to clarify the fuzzy front end of wind turbine innovation process from the RAMS+I point of view as a risk reduction measure at the operational phase.

R. Tiusanen (✉) · J. Jännes
VTT Technical Research Centre of Finland, 1300, 33101 Tampere, Finland
e-mail: risto.tiusanen@vtt.fi

J. P. Liyanage
Centre for Industrial Asset Management (CIAM), University of Stavanger, Stavanger, Norway

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_33,

33.1 Introduction

In the onshore sector, the size of wind turbines and turbine technology seems to have reached an optimum cost/benefit level. Whereas offshore wind turbine technology is still strongly developing to fulfill the specific requirements of offshore operating conditions and high demands of cost reduction by the energy market. The trends in the offshore sector directs toward larger offshore wind farms in the 200–300 MW range and beyond, using dedicated and standardized offshore turbines. According to EWEA, the main focus in the coming years will be on standardizing the installation processes and developing dedicated offshore turbines from a dedicated supply chain, just as it was for onshore wind 15 years ago [1].

Current offshore turbines in shallow waters are mostly developed from onshore designs. According to EWEA's medium and long term scenarios, offshore wind turbine concepts will be changed from onshore based constructions to specifically-for offshore environment designed turbine types. The main driver for offshore wind turbine development is efficiency, rather than generator size. The turbine size is specified by the wind farm operator in the preliminary wind farm design phase based on site specific characteristics and calculations [2]. In the conceptual design phase, turbine manufacturers compare different concepts on how to manage the whole turbine development process: design, manufacturing, assembly, transportation and installation. Examples of important aspects affecting the conceptual design are nacelle concepts and structures, support structures, safe access and maintenance, and offshore logistics. Harsh offshore conditions create new challenges, but it gives also some freedoms to the designers such as esthetics and sound emission level [1].

Wind turbines in cold climates, like in northern Europe, may be exposed to conditions outside the design limits of standard wind turbines. Cold climate-specific issues, such as accessibility, temperature, ice, snow, energy potential, technology, economic risk, public safety, infrastructure, and labor safety, will require additional thought [3].

The market's needs, in offshore wind energy, are contradictory to what comes to turbine development. To achieve large production volumes for the rapidly growing market, the manufacturers should focus on producing continuous, incremental improvements in the current basic concepts to improve product reliability, to increase component lifetime and to develop preventive maintenance strategies. On the other hand, offshore project designers and wind farm operators are requesting completely new—robust, easy to assemble, install and maintain—wind turbine concepts. New designs could be an opportunity to make significant reductions in the cost of energy. In practice, these two strategies will be developed in parallel.

An indicative model for costs of design failures in Dhillon [4] has been presented. According to this model, a fault costing one dollar to repair before publishing the first sketch costs $10 after the publication, $100 in prototyping phase, $1,000 in pre-product phase and $10,000 in production stage [5]. These costs are naturally multiplied in large scale machinery systems, consisting of a large number of machine units, such as wind farms.

In the concept design phase system, designers face the maintenance cost reduction dilemma of improved component reliability versus increased maintenance capability. Because of high reliability requirements and increasing cost reduction demands in offshore wind energy sector, many critical decisions and selections have to be made early in the turbine conceptual design phase. Management of reliability, availability, maintainability and safety issues becomes essential as early as the system requirement specification phase in the beginning of turbine conceptual design phase. Decisions made before the formal development are the most important when thinking about how successful the product finally will become. These decisions strongly define the costs and benefits gained through the whole lifecycle of a product.

In this paper, we present a model for managing RAMS+I factors in the conceptual design phase of offshore wind turbine as a risk mitigating measure under Nordic operational conditions. The model concentrates mostly on technical decisions made in the early development phases. Service aspects, such as maintenance concepts, are not in the focus of this paper. Some level service-based factors still have to be considered in technical decisions made. The objective of the presented model is to clarify the fuzzy front end of the wind turbine innovation process from the RAMS+I point of view.

33.2 Early Phases of Product Development

Various views or models of the product life-cycle exist and can be found, e.g., in EN-50126, EN 60300-1 and EN 60300-2 [5–7], modern design and development literature, and in systems engineering literature and standards. The individual life-cycle phases in the models proposed by various sources typically vary from five to fifteen. The following phases: concept and definition, design and development, manufacturing, installation, operation and maintenance and disposal are most often included in the life-cycle models. These phases are also adopted by, e.g., EN 60300-1 and EN 60300-2 [6, 7]. In this paper we concentrate on describing early phases of a single wind turbine design focusing on the concept and definition phase.

There are many different descriptions of product development. One of the most well-known is the process-model created by Ulrich and Eppinger [8]. Concept development phase is the phase that needs the most guidance in development process. In the Fig. 33.1 there is a description of concept development phase comprised by many different front end activities. Even though the figure shows different stages in sequential form, in real life iteration and back looping is present in all stages [9].

The future is and will always be unpredictable, but with the right measures and techniques, threats and opportunities can be managed. For different kinds of

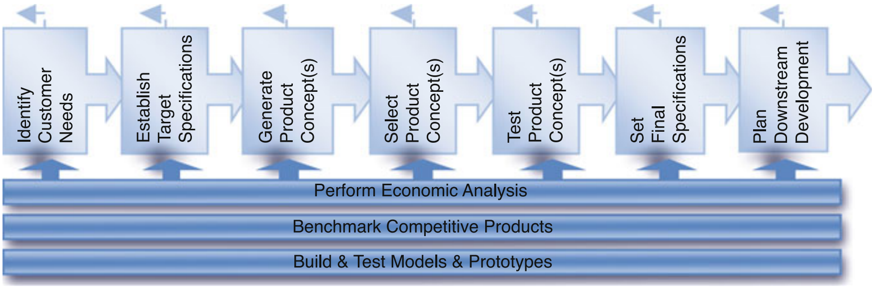


Fig. 33.1 Front end activities [9]

development processes different techniques must be chosen. Early stages of innovation or product development processes are often called the "fuzzy front end" (FFE) [10]. It is the stage that offers one of the biggest opportunities to improve the whole innovation process. At the same time it seems to be the weakest managed part of innovation. Fuzzy front end consists of activities that come before the well-formed new product development process [11]. The term "fuzzy front end" describes the first part of innovation to be a stage dominated by unknown, uncontrollable and unpredictable elements [12]. At the same time it gives an image of something that cannot be managed at any level. It is proposed that the term should be changed to "front end of innovation" so that the people outside the product development branch were not mislead to think that it is just a continued drain of corporate recourses [11]. One of our objectives is to clarify the fuzziness of the early stages of innovation process with systematic RAMS+I management.

Conceptualization is also a term for describing fuzzy front end. Management of front end is a key-factor when creating successful innovations. Front end must be modeled before it can be managed. In practice this means identification of stages, related factors, decision points, specification of incomes and outcomes of each stage and also determination of the information used etc. [10]. The activities which comprise the front end of innovation are decisive whether the product becomes a success or not. Successful companies spend twice as much money and time on these essentials compared to their competitors. Based on the experience, it can be stated that the work done and the resources spent before the actual new product design phase are "refunded" in the later stages of design process [13].

Steps preceding the actual new product development are crucial and define strongly whether the product becomes successful or not. According to Cooper [12] successful companies use approximately twice the time and money to these front end phases compared to their competitors. The following activities can be seen as a part of vital functions preceding new product development:

- Initial screening: the first decision to get into the project.
- Preliminary market assessment: the first and quick market study.
- Preliminary technical assessment: the first and quick technical appraisal of the project.

- The detailed market study, market research, and voice of the customer (VOC) research.
- The business and financial analysis just before the decision to "go to development" [13].

Studies reveal that the work done before the actual development and the resources spent on that point pays for itself in reduced time and improved success rates. There is a common fear of longer development times because of the slowness of the front end activities [13]. Requirements for rapid development somehow conflict with the significance of the decisions made in the concept development phase [14]. The fear for reduced development times is valid, but at the same time experience shows that omitting the front end activities lead to much higher likelihood of product failure. So, in the end there is situation where one has to decide between a possible slightly longer development time and much greater probability of failure [13]. This is one expression of trade-off action vital to product development process.

33.2.1 Stage-Gate® Model

Stage-Gate® is a project management technique, which breaks the innovation process into predetermined stages. Progress from stage to stage is made through gates. Gates control the process and serve as quality control and act as checkpoints where proceeding decisions are made. Every stage is designed so that the information to progress to the next stage is gathered during the stage. Each stage costs more than the preceding one, so the process becomes incrementally committing [15].

In the traditional five-gate model as shown in Fig. 33.2, the first two actual stages are: (1) scoping and (2) building the business case. In proportion, the first three gates are named as follows: (1) idea screen, (2) second screen and (3) go to development. The third gate is the final point where project can be killed before entering the real development where costs grow rapidly [15]. This gate is described to be as particularly critical because at this point it is determined whether the company is ready to invest into the project, and how much [16].

Decisions made in the gates are in the key position when managing risks of product development. Even then these decisions are among the least understood aspects of the product development process. While product development becomes increasingly important for many firms, understanding of gate structures and gate decisions should become better [17]. Also Cooper states that in many companies, the evaluation of projects is weak, ineffective or non-existent [15].

The Stage-Gate® process is an effective tool to accelerate incremental product development. However, it does not directly suit for platform of breakthrough products in the fuzzy front end [11]. Wind turbine development is typically incremental even when offshore technology is in question. Because of the

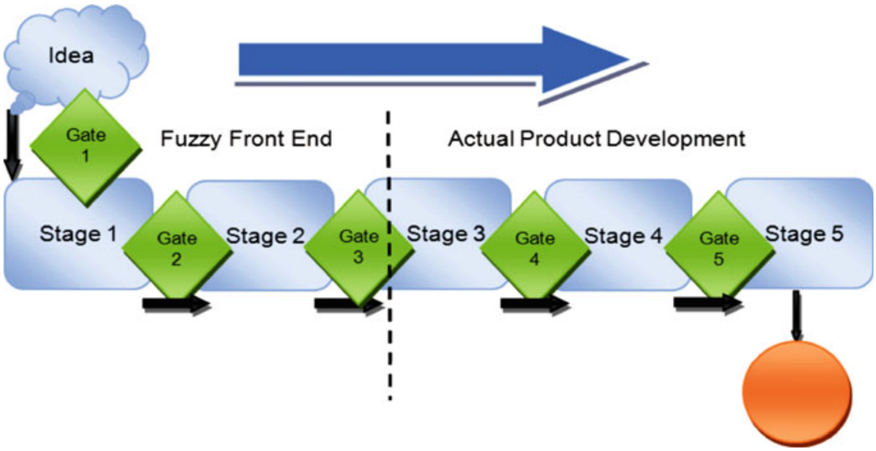


Fig. 33.2 Traditional stage-gate® model (adopted from [15])

applicability to incremental development and the wide use of Stage-Gate® in the development processes, it has been chosen to be one major approach example for the model presented in this paper. Because the Stage-Gate® is widely used in the companies, the model based on it is easier to implement as a part of product development process.

33.2.2 Wind Turbine Conceptual Design

Wind farm design is highly networked in cooperation between wind farm operators, turbine manufacturers, subcontractors, authorities and all other parties involved. Designing an offshore wind farm is an iterative process having several consecutive and concurrent phases. According to EWEA [2] the wind farm design begins with following steps:

- Preliminary design and feasibility study.
- Site investigation and selection.
- Concept development and preliminary layout definition.
- Wind turbine size and model specification and selection.
- Value engineering.
- System specification.
- Wind farm detailed design.

In the wind farm conceptual design phase, the turbine size is specified by the wind farm operator based on site specific characteristics and calculations. The wind turbine is a complex system consisting of various subsystems (e.g., blades, gearbox, generator, brake etc.) which are comprised of numerous components. Wind turbines can also be seen as a subsystem to wind farms which again can be seen as a subsystem to the electric energy supply system and so on. The so-called natural dependability varies largely between different (sub) systems and this should be taken into account when determining and allocating system level RAMS+I requirements at a subsystem level.

Wind farm design process is a sum of work done by different actors. During this process higher level requirements are typically set by the owner and operator of the wind farm and then allocated to turbine supplier and further to subcontractors. RAMS+I management model presented in this paper is built from the turbine supplier point of view, but also subsystem suppliers are able to use the same idea. Early design phases for these actors follow same phases in general. Requirements and their sources vary but the process is somewhat generic. Even though subsystem suppliers develop subsystems from the whole wind turbine point of view, they are developing an independent system from their perspective. Component manufacturers do not need to concentrate on the system availability but only reliability, because components are not designed to be repairable.

Based on observations, it seems that requirements for the next level are made separately in every step. Subsystem suppliers for example do not necessarily know the final application and operation conditions. They only get specifications and requirements for the delivered subsystem. This may lead to problems in specification and requirement definition. Cooper [12] states that two of the worst timewasters in the product development project are constantly changing the project and product definitions. So, it should be recognized that the current way of allocating requirements from actor to another may lead to problems. At the same time the only way of managing requirements when there are so many different sources for requirements is to set them by using the top-down principle [13].

33.3 RAMS+I Tasks in Early Stages

The main message of RAMS+I is that safety and availability issues can be thought and developed concurrently with the main product development process. Usually availability is divided into three parts: reliability, maintainability, maintenance supportability. However, we suggest that the offshore wind area inspectability (I) should be counted as one main factor of availability due to special characteristics of northern offshore wind [18].

When developing a high availability product, it must be made sure that the required performance level is achieved cost efficiently. Replacing less reliable components with highly reliable ones improves systems reliability, but may increase costs [19]. The trade-off between reliability and maintainability is often needed when choosing components for the system under development for availability. Achieving high reliability, especially in complex systems or when system involves untried technology, is very expensive [20].

Analysis of the solution options in the beginning of the product development phase creates possibilities to identify problems related to availability and safety.

With the help of the analysis results, it is possible to create solutions for resolving recognized problems. Analyzing the group and its members, they should have all the needed information on operation and maintenance concerning existing similar systems on the market. Performing analysis in the beginning of product development process makes it possible to achieve remarkable innovation to satisfy customer needs [21]. RAMS+I characteristics should be taken into account in the design phase to be able to manage business risks associated with non-performance of the system.

The RAMS+I process is a way to manage safety and availability related requirements appropriately straight from the beginning of product development. The most important decisions concerning product safety and availability are made already in the beginning of the lifecycle. The purpose for the RAMS+I process is to consider the safety point of view in parallel with availability one. The examination is made with the whole lifecycle of a product in mind. Outputs from the previous stage serve as inputs to the next one. Concurrent engineering is a work methodology based on the parallelization of tasks (i.e., performing tasks concurrently) [22]. It refers to an approach used in product development in which functions of design engineering, manufacturing engineering and other functions are integrated to reduce the elapsed time required to bring a new product to the market.

One availability management problem is that availability aspects are often being considered when the product is already on the market. The challenge is to make availability an integrated part of product development and to make availability a built-in feature [21].

33.3.1 Up-Time—Down-Time

Figure 33.3 presents availability-related time concepts. The figure is made from the system point of view, which in this case means that such external factors that cannot be influenced in the product development phase are removed. These external factors include things like strikes, economic situations, wind condition related shutdowns, or shutdowns caused by other external reasons not exactly availability-inspired. The scaling of the boxes in figure does not reflect duration of different phases but is driven by drawing technical reasons.

There are various definitions for availability measurement. Availability can simply be defined as in Eq. 33.1, this classical definition is presented in various sources (e.g., IEC 60050(191) [23]).

$$A = \frac{T_{up}}{T_{up} + T_{down}} = \frac{MTBF}{MTBF + MTTR} \tag{33.1}$$

where A is availability, T_{up} is a time when system is available and T_{down} is time when system is unavailable due to corrective maintenance. This definition for availability takes into account only shutdowns caused by corrective maintenance. Equation must be extended with MPMT if there is a need for taking preventive maintenance causing shutdowns into account. It must be pointed out that not all the preventive maintenance actions require shutting down the system, and therefore lower availability of the system.

According to Fig. 33.3 and the definition of availability, affecting any of the earlier defined parts of availability can be a way to improve the total availability of the system. For example, time to failure can be affected straight with reliability, improving maintainability, inspectability or maintenance supportability time to restoration can be affected.

In Nordic conditions, special attention must be paid to waiting time. Harsh conditions such as ice and snow during a long winter affects waiting times, while there are certain periods when the site is inaccessible. Ice and cold climate in its entity also make maintenance more difficult and time consuming, therefore extending repair and preventive maintenance times. Hard conditions set new requirements for component reliability. It is strongly suggested that both access methods and the component suitability for Nordic sites are considered.

33.3.2 LCC—LCP

Cost predictions and related calculations are closely connected to the RAMS+I study. Both the maintenance costs related to different fault situations and the production losses are evaluated through cost analyzes. Combining cost factors from RAM analyzes with the lifecycle costs of the whole system creates the

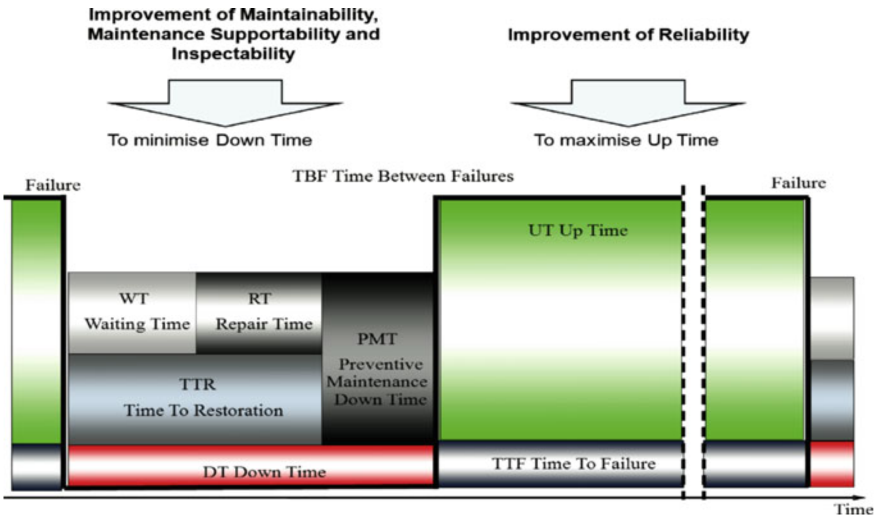


Fig. 33.3 Availability factors and *up/down* times

possibility of performing a comparison between solution options and the ability to decide which of the option would be the most reasonable to carry through. “The costs of unreliability in service should be evaluated early in the development phase, so that the effort on reliability can be justified and requirements can be set, related to expected costs” [20].

The RAMS+I process gives good possibilities to plan and optimize the maintenance of the systems. It improves the ability to plan the maintenance services and the right pricing. If the operating costs are not estimated precisely enough, great risks are taken when signing maintenance agreements. In addition to the life-cycle costs, it is important to notice also the profits and cost savings related to the better availability performance.

Cost-benefit estimations and calculations based on the estimations are closely connected to RAMS+I process. Combining those calculations of life-cycle-based availability costs to life-cycle cost of a whole product gives an opportunity to compare different solutions and to make the right decision of which solution would be the best to execute. It should also be pointed out that the aim of RAMS+I management should not only be to lower life-cycle costs. Cost savings and earnings due to improved availability and safety should also be estimated [22].

As can be seen in Fig. 33.4, system availability management creates the possibility to predict O&M costs. This of course requires reliable data or successful forecasts based on the available information. Systematic analysis offers this valuable information when developing totally new products or when transferring existing technology to totally new operational environment, as is the case when moving wind technology from onshore to challenging offshore environment.

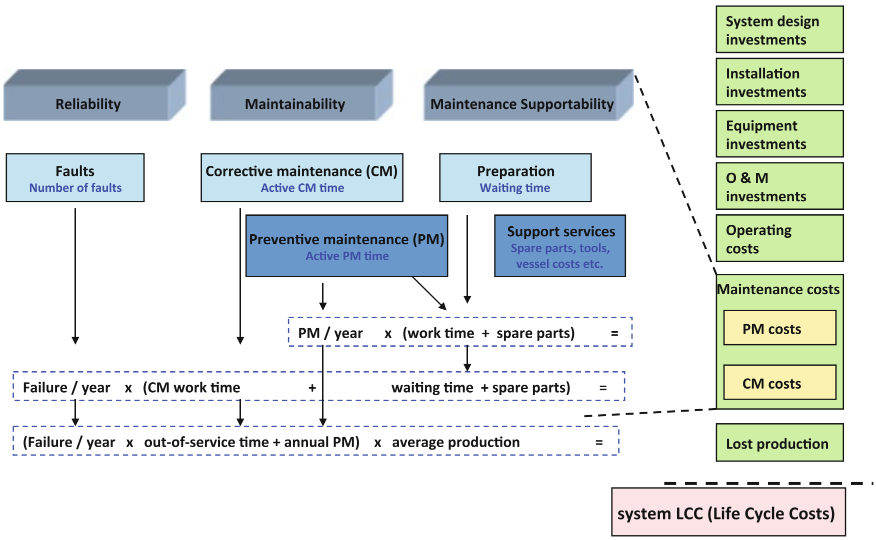


Fig. 33.4 Simplified life-cycle-cost model and costs related to system availability

In the Nordic sea areas, site conditions impact strongly on the calculations. Therefore, all the cost-benefit calculations must be done site-specific. Nordic conditions such as ice, snow, operation and maintenance temperatures, wave height etc. must be evaluated exactly especially when no real-life data from similar sites in not available.

33.3.3 Risk Analysis

RAMS+I information needed to support development can be reached already in the early phases of the project by carrying out systematic analyzes. These analyzes become more detailed while the development project proceeds; however, in early phases, analyzes are committed in a very general level. In new concept development projects, the available system information in the early phases is typically limited and inaccurate. Qualitative RAMS+I analysis methods can then be used to identify and analyze the most important issues on a rough level. As the project proceeds more detailed information is available and analysis and assessments can be done in more detailed level. Most of the analysis methods can be carried out for different accuracy levels. For example, FMECA analysis can be executed at a functional level instead of component level as usual.

Models presented in this paper contain proposals for RAMS+I based risk analyzes methods suitable for early stages of wind turbine development. Theses analysis methods are chosen so that they would support the decisions in the beginning of turbine development and would make availability and safety management possible in the front end process. Analyzes are selected keeping the amount and quality of information in mind, also the characteristics of wind turbines are considered. There has also been the intent to cover the human–machine interaction as extensively as possible.

33.3.4 Documentation and Data Historians

There are numerous benefits of the systematic and well-documented RAMS+I process. Practical actions to fulfill RAMS+I requirements can easily be demonstrated to the customer with the help of documents created during the process. Documentation can also be used while selling the product to justify a higher purchase price and demonstrate cost savings and benefits gained during the whole lifecycle of the product.

Documentation is a vital part of risk analyzes. Documentation must be exact so that the findings can be traced back and reviewed later. Identifying and analyzing risks is not enough, also the measures to be taken to manage risks must be decided. Analyzes must be updated whenever necessary.

33.4 RAMS Management Model for Offshore Wind Turbine Concept Development

Figure 33.5 presents front end phases that are related to the wind turbine design process. The model presented in this paper is based on different process models. The presented model is strongly influenced, for example, by Cooper's Stage-Gate® model and Ullrich and Eppinger's product development description. The main idea of the model is to divide the front end process into small and easily controllable phases and gates, where decisions concerning the future of the idea under development are made based on the criteria defined. Decisions that can be given to the idea on the gates are "*go*", "*hold*" or "*kill*". *Go* means that development of the idea can move to next step or stage for further development *kill* decision means that development of the idea stops. *Hold* decision is given to the idea that has potential but for some reason cannot be realized, for example because of lack of requisite technology. Hold can mean either putting the idea to wait for better times or sending it back to previous development stage where it can be improved to meet the requirements of a gate.

Although the model represents product development as a somewhat straightforward process, in real life iteration is present all the time. For all practical purposes, iteration is needed to achieve the best result, although it might not be the most effective way to proceed.

Making changes or bad gate decisions becomes more expensive, when moving from previous step to the next one. This is the main reason why proceeding decisions in the gates must be done critically based on the tough criteria. Gatekeepers should be separated from those who are involved in development process to guarantee the objectivity of the decisions [16]. The decision to go to the turbine development process must be especially carefully thought through.

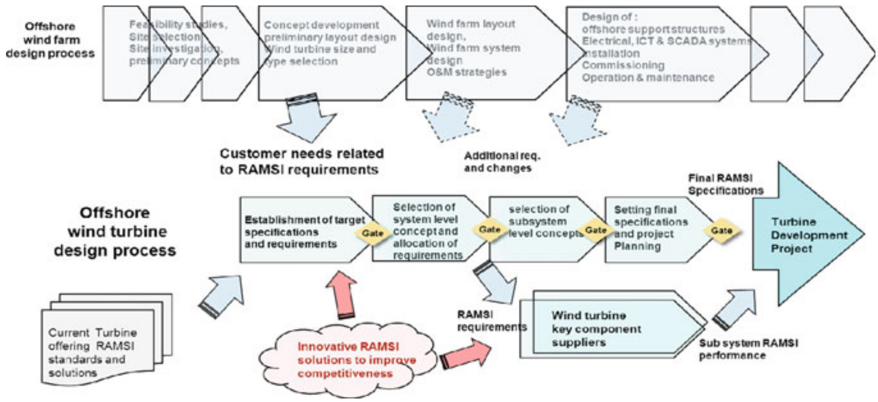


Fig. 33.5 Simplified RAMS+I information flow in the early phases of wind turbine design

Table 33.1 RAMS+I management for wind turbine development in early phases

Stage	RAMS+I objectives	Inputs	Tasks	Examples of analysis methods	Outputs
Identification of customer needs and existing solutions	Identification of accident hazard related to ideas	Ideas to be developed	Establishing of RAMSI criteria and requirement list on system level	Potential problem analysis	Preliminary hazard list
	Identification of RAMSI aspects	Customer needs	Outlining differences between present and target state	RAMSI brainstorm	Preliminary requirement list for the whole system RAMSI requirements
		Existing solutions and current turbine offering			Requirement list for the whole system
		Experiences from existing similar devices (qualitative and quantitative)			
Establishment of target specifications and requirements	Establishment of general availability and safety requirements	Preliminary list of system level RAMSI requirements	Clarifying the sources of requirements	Checklist	RAMSI specifications and requirements in system level
	Verifying the treatment of RAMSI aspects	Preliminary hazard list for qualified product ideas	Setting RAMSI requirements in system level		Supplemented requirement list for the whole system
		Special requirements set by conditions of use	Consideration of possibilities to fulfill availability requirements		Documented decisions of killed and held ideas
		Standards and orders of authorities			

(continued)

Table 33.1 (continued)

Stage	RAMS+I objectives	Inputs	Tasks	Examples of analysis methods	Outputs
Selection of system level concept and allocation of requirements	Identification of hazards related to different concepts	RAMSI specifications and requirements for the whole system	Creating hazard lists for concepts	PHA	Hazard list for system level concepts
	Identification of subsystems and allocation of RAMSI requirements	System level concept alternatives	Identification of subsystems	FMECA (on function level)	RAMSI requirements in subsystem level
	Risk evaluation for concept alternatives	Business model for the product (device delivery vs. lifecycle service providing)	Criticality analysis for subsystems	Cost-benefit analysis	Criticality evaluation for subsystems
		Requirement list	Evaluation of natural dependability of subsystems		Allocated subsystem requirements
		Hazard lists	Setting RAMSI requirements for subsystems		System level concept
		System level concept alternatives	Consideration of availability and its elements for the product		Document of held and abandoned concepts
			Comparing solutions with cost-benefit analysis		

Table 33.1 (continued)

Stage	RAMS+I objectives	Inputs	Tasks	Examples of analysis methods	Outputs
Generation and selection of subsystem level concepts	Identification of accident hazards and hazards related to availability	Subsystem specific RAMSI requirements	Methods of implementations for identified subsystems	Work safety analysis	Document of identified accident hazards and hazards related to availability
	Choosing of the most optimal subsystems from the RAMSI point of view	Criticality evaluation for subsystems	Ensuring compatibility of subsystems	FMECA (for subsystem functions)	Document of risks related to use of new technology
		Allocated subsystem requirements	Identification of risks related to possible new technologies in the product		Document of abandoned concepts
		System level concept			Subsystem level concept

Table 33.1 (continued)

Stage	RAMS+I objectives	Inputs	Tasks	Examples of analysis methods	Outputs
Setting final specifications and planning of the development project	Verification of achievement of RAMSI requirements	Subsystem level concept	Identification of tasks and related error possibilities	Fault trees (qualitative)	Final RAMSI specifications to support product development
			Consideration of maintenance points of view	HRA (first two steps)	Hazard list
				Maintenance checklist	Reports of the analysis done
					Systems' operation and maintenance concept
					List of modifications related to requirements
					RAMSI plan
					Plan for the continuation of development project
					Decision of starting a development process

In Table 33.1, a RAMS+I management “flow” is described. Table shows RAMS+I objectives, inputs, general tasks, analysis method recommendations and outputs for every identified wind turbine development stage in the front end (see Fig. 33.5).

33.5 Conclusion

RAMS+I requirements are meant to guide the system development process in its various phases. The exploitation of a systematic RAMS+I program aims to identify, analyze and assess availability and safety issues and specify requirements in the most appropriate way. RAMS+I requirement management, as an essential part of the general systems engineering approach, tries to help avoid situations where defects in availability and safety performance are not detected until the system is already operating. Corrective actions are then difficult and expensive to execute.

In this paper, we have outlined the early stages of wind turbine design and RAMS+I management tasks related to these stages. The presented RAMS+I model is generic for the offshore wind sector and it might not be applicable as it is not for any company. More research is needed to build company-specific models based on the model presented in this paper.

Because of the fast development of offshore technology and the introduction of new innovative turbine concepts, quantitative analysis methods cannot be used. Relevant data is not available. Conceptual design must be supported by qualitative risk analysis and reliability analysis methods.

RAMS+I related lifecycle cost-benefit evaluations benefit both suppliers and customers. This message should be made clearer for both directions to avoid situations where the customer only stares at the purchase price. To avoid this, wind turbine manufacturers must be able to provide evidence of the costs and benefits of the whole lifecycle of their product more clearly and precisely. The presented model gives a good start for this kind of development.

Further research is needed for case specific model implementation and verification. One main interest in RAMS+I management development is information transfer between different actors in the supply chain.

Acknowledgments This work has been (partially) funded by Norwegian Center for Offshore Wind Energy (NORCOWE) under grant 193821/S60 from Research Council of Norway (RCN). NORCOWE is a consortium with partners from industry and science, hosted by Christian Michelsen Research.

References

1. European Wind Energy Association (2009) Oceans of Opportunity. Harnessing Europe's largest domestic energy resource. European Wind Energy Association, September 2009
2. European Wind Energy Association. (2009) Wind Energy—the facts. European Wind Energy Association, Available on website:www.windfacts.eu
3. Baring-Gould I, Tallhaug L, Ronsten G, Horbaty R, Cattin R, Laakso T, Durstewitz M, Lacroix A, Peltola E, Wallenius T (2010) Recommendations for wind energy projects in cold climates. VTT working papers 151:61
4. Dhillon BS (1999) Design reliability—fundamentals and applications. CRC Press LLC
5. EN-50126 (1999) Railway applications—the specification and demonstration of reliability, availability, maintainability and safety (RAMS). European Committee for Electrotechnical Standardization—CENELEC, p 76
6. EN 60300-1 (2003) Dependability management. Part 1: dependability management systems. International Electrotechnical Commission (IEC), Geneva
7. EN 60300-2 (2004) Dependability management. Part 2: guidelines for dependability management. International Electrotechnical Commission (IEC), Geneva
8. Ulrich KT, Eppinger SD (2004) Product design and development, 3rd edn. McGraw-Hill, New York
9. Paasi J, Luoma T, Strong R, Zhou R (2008) Systematic strategic decision support for innovation development. In: Proceedings of European conference on management of technology and innovation (EuroMOT), Nice, France
10. Koen P, Ajamian G, Burkart R, Clamen A, Davidson J, D'Amore R, Elkins C, Herald K, Incorvia M, Johnson A, Karol R, Seibert R, Slavejkov A, Wagner K (2001) Providing clarity and a common language to the "fuzzy front end". Res Technol Manage 44(2):46–55
11. Koen PA, Ajamian GM, Boyce S, Clamen A, Fishe E, Fountoulakis S, Johnson A, Puri P, Seibert R (2002) Fuzzy front end: effective methods, tools, and techniques. In: Belliveau P, Griffin A, Somermeyer S (eds) The PDMA toolbook for new product development 1. Wiley, New York, pp 5–35
12. Cooper CG (2005) New products—what separates the winners from the losers and what drives success. In: Kahn KB (ed) The PDMA handbook of new product and development, 2nd edn. Wiley, Hoboken, pp 3–28
13. Apilo T, Taskinen T (2006) VTT tiedotteita 2330: innovaatioiden johtaminen. Espoo, Otamedia Oy, p 62 (In Finnish)
14. Cooper RG (2001) Winning at new products: accelerating the process from idea to launch, 3rd edn. Basic Books, New York, p 425
15. Kim J, Wilemon D (2002) Focusing the fuzzy front-end in new product development. R&D Manage 32:269–279
16. Schmidt JB (2005) Gate decisions: the key to manage risk during new product development. In: Kahn KB (ed) The PDMA handbook of new product development, 2nd edn. Wiley, Hoboken, pp 337–348
17. Koen PA (2005) The fuzzy front end for incremental, platform, and breakthrough products. In: Kahn KB (ed) The PDMA handbook of new product development, 2nd edn. Wiley, Hoboken, pp 81–91
18. Tiusanen R, Jännes J, Liyanage JP (2011) RAMSI management—from single analyzes to systematic approach. In: Proceedings of the 24th international congress on condition monitoring and diagnostics engineering management (COMADEM). Stavanger, Norway
19. O'Connor PD (2002) Practical reliability engineering, 4th edn. Wiley, Chichester, p 513
20. Virtanen S, Hagmark PE (1997) Käyttövarmuus tuotesuunnittelussa—ratkaisujen etsiminen ja valinta, Julkaisu nro B 21. Espoo, Teknillinen korkeakoulu, Konetekniikan osasto, p 60 (In Finnish)

21. Markeset T, Kumar U (2003) Integration of RAMS a risk analysis in product design and development work processes—a case study. J Qual Maintenance Eng 9(4):393–410
22. Kivipuro M, Reunanen M, Valkokari P (2008) Hallitse RAMS-vaatimukset. Promaint 7:27–29 (in finnish)
23. IEC 60050(191) (1999) Electro technical vocabulary. Dependability and quality of service. IEC

Chapter 34
Success Factors for Remote Service Systems

G. Schuh, C. P. Winter, C. Grefrath and P. Jussen

Abstract Remote services are services enabled by information and communication components and therefore do not require the physical presence of a service technician at the service object to provide a task. The impact of remote service on the capital goods industry has been increasingly significant over the recent years. Still many companies struggle with developing and implementing successful business models for remote service. This leads to a lot of unaccomplished benefits for the customer as well as for the companies themselves. A survey throughout companies in the industrial machine and plant production sector was conducted in order to determine what successful companies do differently from those that cannot effectively implement remote service business models. The study presented in this chapter identifies key success factors of companies that effectively implemented remote services for their products. In order to identify the successful companies a scale for measuring remote service success was developed. Only by the use of this scale further findings regarding the success factors were possible. Key findings include the fact that successful companies actively market their remote service to their customers. Generally they try to approach their remote service business from the operating company's perspective.

G. Schuh (✉) · C. P. Winter · C. Grefrath · P. Jussen
FIR (Research Institute for Industrial Management) at the RWTH Aachen University, Pontdriesch 14/16, 52062 Aachen, Germany
e-mail: schuh@rektorat.rwth-aachen.de

C. P. Winter
e-mail: cord-philipp.winter@fir.rwth-aachen.de

C. Grefrath
e-mail: christian.grefrath@fir.rwth-aachen.de

P. Jussen
e-mail: philipp.jussen@fir.rwth-aachen.de

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_34,

34.1 Introduction and Initial Situation

The technical aspect of remote service has been covered extensively by research, producing a variety of technical solutions for almost any problem. It is estimated that the use of remote service can reduce after sales costs by as much as 30 % [1, 2]. However, especially small and midsize companies struggle in successfully adding remote service to their product portfolio. Scientific research and literature has identified several barriers that result in companies not accomplishing the full potential of remote service. Providers of remote service often find it difficult to specifically identify and address the true use for their customers. In addition visibility of remote services for the customer is limited. This leads to in-transparency for the pricing process. The customer is not aware of how big the service provider's effort really is resulting in a decreasing willingness to pay for the service. The lack of awareness for the remote service's potential on the customer side is closely linked to insufficient and ineffective marketing strategies on the provider side. Also, for many customers, there remain significant uncertainties considering the transmitted data and its confidentiality [3, 4]. To address the practical relevance of the matter this chapter widens the focus in order to capture the development of remote service business models including marketing.

A practical approach on how to measure successful implementation of remote service system is missing and therefore possibilities for reliable identifications of success factors were limited. Before further considering the success factors a scale for measuring the successful implementation of remote service had to be developed first. It functions as the basis on which the subsequent research was conducted. The aim of the research was to find out:

- Whether it is possible to develop a reliable scale to measure successful implementation of remote service systems?
- What distinguishes successful from less successful companies?

The first point will be thoroughly discussed in Sect. 34.3.2. For the second point, a few theoretic deliberations need to be made. In order to evaluate and point out success factors for remote service implementations, the business model aspect of remote services had to be analyzed. Three elements of the business models are taken into account when designing the survey:

- **The range of service products and the market approach**: This part of the business model defines the allocation of different products and services to different markets. New services influence the existing portfolio of a company. A company which introduces new services to their portfolio has to make considerations on how to fit those services into the existing portfolio. Generally a structured approach is considered to be of most success [5]. Defining a target customer group and addressing it by creating specific customer value is a major part of creating new customer potential [6]. In the case of the already existing customer the company has to focus on tying the customer further to the company. Part of the marketing approach is also the sales and distribution strategy.

Although not commonly used, incentive schemes are considered to be highly relevant for implementing services [7]. Most importantly, a consistent orientation toward the customer benefit exhibits a basis for the success of the service in a market environment [8, 9].

- **Production of the services**: This stage defines how the service is provided. The allocation of resources, make-or-buy decisions, and the allocation of tasks for value creating partners are included in this part of the business model. The focus in the production of services is on the development and organization of required processes as well as required resources, technologies, and employees. As a first step the companies' disposable resources have to be analyzed to assure that newly developed service products only use available resources. Typically services include a distinct process component. The value chain concept enables companies to systematically visualize and develop services with a high orientation toward adding value for the customer [10–13].
- **Profit mechanism**: Companies have to consciously determine a pricing strategy and create a basis for profits. The integration of services and tangible assets is an increasing factor for business success [14]. The customer perception of the price is also to be taken into account [15]. Just as for products and services there also exists a lifecycle for prices and revenue models.

34.2 Research Design

To tackle the research questions we set up an online survey, designed to give answers in three main areas:

- Demographic data (company size etc.)
- Measurement of successful remote service system
- Success Factors

The first part of the survey focuses on general data describing the companies' key features. The question includes facts like company size, revenue, sector, etc. Also the target group of their main products was determined as well as which services were specifically offered. The second part of the survey aims at measuring the success of the companies' remote service portfolio. The success of a company was measured by using different statements. On a scale from 1 to 6 the participants were asked to state their level of agreement with each single statement. By using factor analysis the results of the statements were aggregated in order to gain a final classification and distinguish between successful and not successful companies. The third part of the survey considers the practical design of the remote service in each company. It addresses the second aim of the research by looking into the success factors of the business models. In this part, the questions are based on the theoretical fundamentals regarding business models as briefly described in Sect. 34.1.

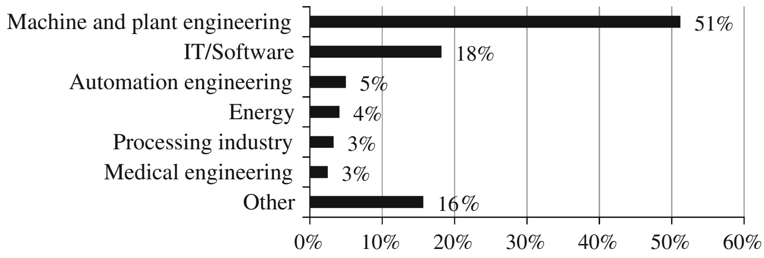


Fig. 34.1 Surveyed companies by sector

34.3 Research Results

34.3.1 Demographic Data

A total of 122 companies answered the survey. Their demographic data is summarized below. The following figure shows how the companies are divided up into different sectors. 51 % of the surveyed companies are based in the machine and plant engineering sector. The second biggest single sector is the IT/software sector. In general it can be said, that the survey focused on the machine and plant engineering sector (Fig. 34.1).

Furthermore a big part of the companies are considered to be of medium size as the following figures illustrate. Over half of the participating companies had more than 1000 employees. In terms of revenue the sample group was almost equally divided into companies with more than 125 million euros and companies with less than 125 million euros (Figs. 34.2, 34.3).

34.3.2 Measurement of Remote Service Success

In order to measure success in the implementation of a remote service system, a new measurement scale had to be introduced: As items five statements are used with an answer scale ranging from 1 to 6 and participants were asked to assess the status of their company in regard to these statements. Each time a high level of

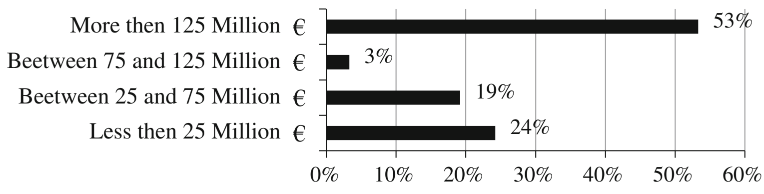


Fig. 34.2 Surveyed companies by total revenue

Fig. 34.3 Surveyed companies by number of employees

agreement with a statement (5 or 6) would represent a high level of success whereas a low level of agreement would mean that the company has not yet been able to successfully use remote service. The following statements were used: [*Note*: the original survey was conducted in German, therefore a translated version of the items is described here. The English items were not tested in a survey context].

"We achieved to…

- …successfully bring a remote service system to market." (item1)
- …implement remote services as an independent product." (item2)
- …independently price remote services." (item3)
- …understand and implement customer needs for remote service." (item4)
- …incentivize the use of remote services." (item5)

Using a factor analyses (Principal Component Analysis, conducted in SPSS 18) a single factor with high factor loads was withdrawn from the five items, showing the applicability of the scale. The SPSS outtakes "component matrix" and "KMO and Bartlett's test" are shown in the figure below. Additionally, the reliability tests revealed a high Cronbachs alpha of 882 for the 5 items, showing that the developed scale is indeed usable to measure remote service success (Fig. 34.4).

Building on that factor we were able to conduct a hierarchical cluster analysis that revealed two clusters. The first cluster contained the companies which rated to

Component Matrix[a]

	Component
	1
ITEM 2	,875
ITEM 1	,843
ITEM 5	,813
ITEM 3	,807
ITEM 4	,793

Extraction method:
Principal Component Analysis

a.1 Components Extracted

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		,799
Bartlett's Test of Sphericity	Approx. Chi-Square	349,381
	df	10
	Sgi.	,000

Fig. 34.4 SPSS component matrix and KMO and Bartlett's test

be successful in operating a remote service business model. The second cluster included the companies which rated to not be successful in doing so. Overall about 2/3 of the surveyed companies were allocated in the first cluster, whereas the rest of the companies fit into the second cluster.

As evaluation the scale was referenced against the basic economic data given by the each of the companies as well as against the question whether the remote service system was profitable. The evaluation revealed that the scale did indeed measure successful implementation of remote service systems.

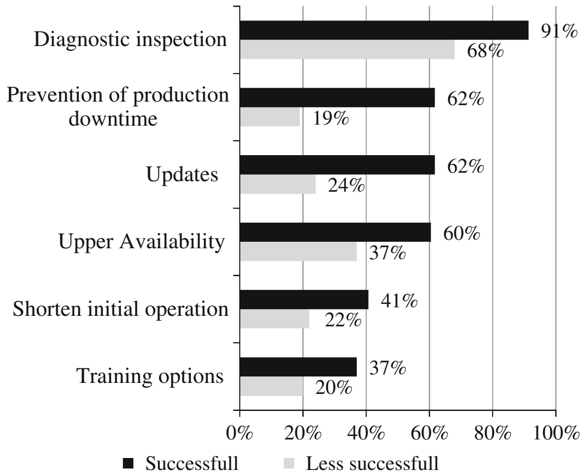
34.3.3 Success Factors for Remote Services

In the following section key findings are presented to illustrate the results that were gathered from the survey. As a general statement the survey showed a higher customer orientation in successful companies. They shaped their remote service portfolio specifically to create customer value. Figure 34.5 shows that successful companies include a wider range of key services in their portfolio.

Successful companies strive for servicing a wider market by caring for a bigger percentage of the installed base (although there is still room for improvement) and focusing also on third party and old equipment. This shows a better market awareness and implementation of the customers' needs into the portfolio which leads to a bigger market potential for the successful companies (Figs. 34.6, 34.7).

Taking a more specific look at the remote service business model, the companies were asked about the use of the data which they gathered in their remote service activity. Figure 34.8 shows that successful companies frequently use the data not only for their own benefit but also for the benefit of the customer. Successful companies therefore not only position themselves with regard to

Fig. 34.5 Which demands are addressed in the remote service portfolio?



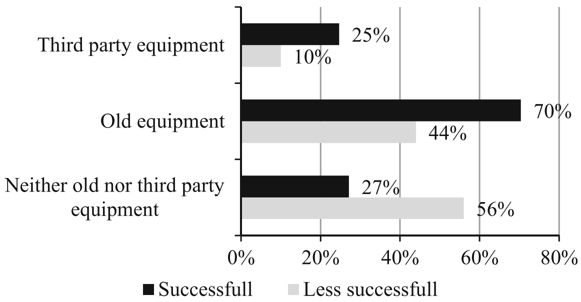


Fig. 34.6 Remote service for old and third party equipment

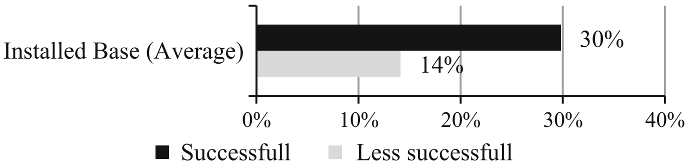


Fig. 34.7 Installed base

customer needs on a wide market scale, but also specifically design detailed aspects of their service products to create customer value.

Another key finding contributes to the fact that successful companies have a much better lifecycle orientation in their remote service systems. They tend not to focus on the first stages of their products lifecycle but rather address and implement their services for the later stages, where most customer benefits and therefore the highest willingness to pay exist (Fig. 34.9).

Success of remote service systems also depends on the promotion and marketing a company is using to push their service product toward their customers. The survey showed that successful companies use a wider range of promotional

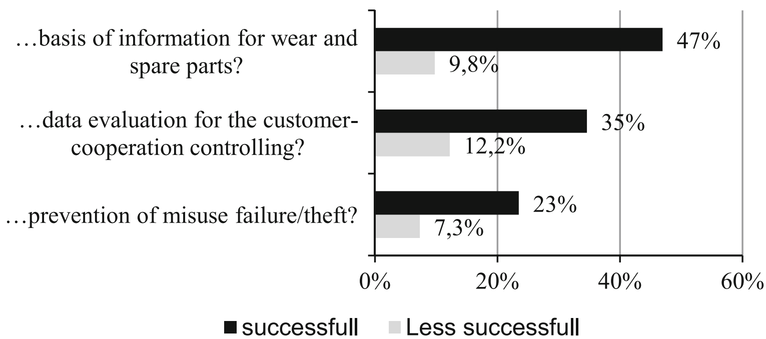


Fig. 34.8 Use of the data gathered through remote service

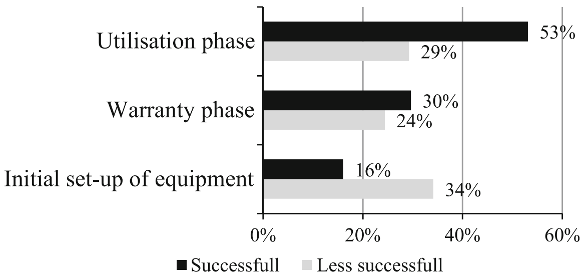


Fig. 34.9 Remote service in different lifecycle stages

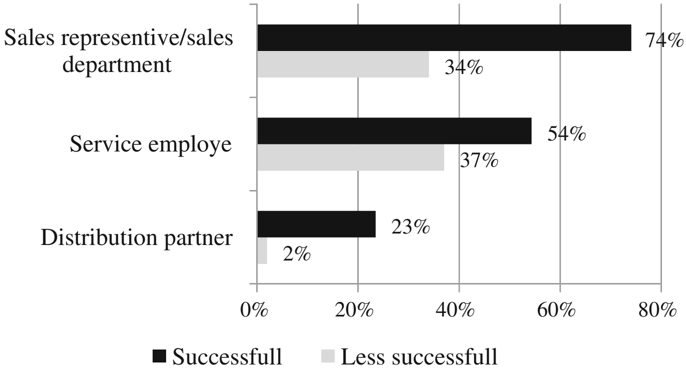


Fig. 34.10 Promotion of remote service

channels. They market their services not only through their sales departments but also via their service employees and with the help of their distribution partners (Fig. 34.10).

34.4 Conclusions

This chapter contributes two important aspects to the research on remote service systems. First a new scale for measuring the success of the implementation of remote service in companies was introduced and verified. It is based on a self-assessment with the help of five different statements and allows for companies as well as for future research to put the success of a remote service business model into perspective.

Secondly key success factors for the implementation of remote service business models were identified in the study. The success factors presented in this chapter can be summarized by three major statements:

- Companies that have successfully implemented remote service in their product portfolio have shown a high orientation toward customer needs and toward creating customer value.
- Companies that have successfully implemented remote service in their product portfolio focus on the whole product lifecycle when developing and offering their services.
- Companies that have successfully implemented remote service in their product portfolio and promote and market their service products actively to their customers.

References

1. Böhle F, Pfeiffer S, Sevsay-Tegelhoff N (2004) Die Bewältigung des Unplanbaren, Verlag für Sozialwissenschaften, München
2. Anderson JC, Narus JA (1995) Capturing the value of supplementary services. Harvard Bus Rev 73(1):75–83
3. Rühmann N, Schröder T (2005) Umfang und Nutzung von Serviceleistungen in der Kunststoff verarbeitenden Industrie. Studie Im Rahmen des Forschungsprojektes “T-SMS”, Forschungsinstitut für Rationalisierung, Aachen
4. Wünderlich NV (2009) Acceptance of remote services: perception, adoption, and continued usage in organizational settings. Gabler, Wiesbaden
5. Schuh G, Schmitt R (2010) Geschäftsmodelle für technologieunterstütze, produktionsnahe Dienstleistungen des Werkzeug- und Formenbaus. Apprimus Verlag, Aachen
6. Müller-Stewens G, Lechner S (2005) Strategisches Management. Wie strategische Initiativen zum Wandel führen. Schäffer-Poeschel, Stuttgart
7. Gudergan G, Thomassen P, Kutlina Z (2007) Fakten und Trends im Service. Aachen
8. Boyt T, Harvey M (1997) Classification of industrial services: a model with strategic implications. Ind Mark Manage 26(4):291–300
9. Porter M (2008) The five competitive forces that shape strategy. Harvard Bus Rev 01:78–93
10. Johnson M, Christensen C, Kagermann H (2008) Reinventing your business model. Harvard Bus Rev 86(12)
11. Benkenstein M (2002) Strategisches marketing, p. 101ff
12. Spiegel T (2003) Prozessanalyse in Dienstleistungsunternehmen, P. 32ff
13. Allee V (2000) Reconfiguring the value network. J Bus Strategy 21(4):36–41
14. Schuh G, Georgi L (2008) Kundenorientierte Konfiguration von Leistungsbündeln. Sales und Service—Management, Marketing, Promotion und Performance. Hrsg.: Keuper, Frank; Hogenschurz, Bernhard. Gabler, Wiesbaden, pp 61–91
15. Diller H (2008) Preispolitik. Kohlhammer, Stuttgart

Chapter 35
Customer Observation as a Source of Latent Customer Needs and Radical New Ideas for Product-Service Systems

J. Hanski, M. Reunanen, S. Kunttu, E. Karppi, M. Lintala and H. Nieminen

Abstract The importance of maintaining close contact with customers and utilizing customer-based information has been emphasized in the industrial service and product-service systems (PSS) literature. A profound understanding of the customer's business and production environment is needed for successful PSS development. The conventional methods for gathering information about customers (surveys, feedback and interviews) typically result in incremental improvements and information about existing products and services. The focus of this paper is on how the information and ideas from customer contacts can be better captured to enable radical improvements. A framework for capturing the customer ideas is presented. The framework is based on customer observation methodology, entrepreneurial opportunity recognition model, front end of innovation literature as well as the experiences of a case study and interviews.

35.1 Introduction

The open innovation paradigm [1] states that the utilization of both the internal and external ideas and knowledge are important for the success of the company. The value of the customer as a source of knowledge and ideas is emphasized in [2] and [3].

J. Hanski (✉) · M. Reunanen · S. Kunttu
VTT Technical Research Centre of Finland, PO Box 1300 FI-33101 Tampere, Finland
e-mail: jyri.hanski@vtt.fi

M. Reunanen
e-mail: markku.reunanen@vtt.fi

S. Kunttu
e-mail: susanna.kunttu@vtt.fi

E. Karppi · M. Lintala · H. Nieminen
Fastems Oy Ab, Tuotekatu 4 FI-33840 Tampere, Finland
e-mail: esa.karppi@fastems.com

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_35,

Being close to the customer provides access to the customer's needs and problems [4]. However, the benefits of customer-based knowledge have not yet been fully utilized. This is partly due to the lack of tools for effectively involving the customer in the knowledge creation processes and partly to the fact that customer is seen as a source of imitative and incremental improvements [3, 5]. The conventional marketing research tools, such as interviews, focus groups and surveys, focus on understanding the expressed needs of customers and lead to incremental improvements of products or services [6].

For this paper, interviews were conducted with eight Finnish business-to-business companies operating mainly in the manufacturing industry. The questions asked concerned the identification of customer needs and customer based information.

According to the interviewees, they strive to systematically collect use-based information from their customers. This information is in the form of feedback on faults, product use and customer satisfaction. For a more comprehensive picture of how the products and services are used, more diversified information might be needed. The use-based information is also frequently scattered over many systems and databases, which limits the accessibility of the information.

Identifying present and future customer needs is a task of great importance at the front end of innovation. The latent customer needs are especially important as they are a source of radical innovations. The companies are aware of the difficulties in defining future customer needs, but do not provide effective solutions to this problem. Instead of latent needs the emphasis of the companies is on the needs the customers are able to articulate. Some of the interviewees stressed the importance of being close to the customer and in this way finding out the customer's real needs. This is also a good basis for finding out the latent needs, but may not suffice alone. Finding out the latent needs effectively may require a specific methodology.

In most of the companies, the identification of customer needs is in the hands of the sales department, which works at the customer interface. This is encouraging because of the good perspective these employees have on customers' needs. The companies agree on the inclusion of a customer in the development processes of product and service at an early phase. The ways of including the customer involve discussions on potential new products and services, and testing new ideas and products. This leaves ample opportunity for further development in utilizing the customer in the development processes. For instance, observing the customer's use of products and thus evincing new customer-oriented ideas are proposed in this paper.

The implementation of service contracts has increased the amount of information available on products and services after delivery. The customer based information is mostly collected by surveys, remote diagnostics and maintenance reports. Nevertheless there is a need to ascertain how the products are used in reality and how they meet the needs of the customers. There are systematic methods for analyzing and utilizing this customer based knowledge in half of the companies represented by the interviewees. Most of the methods concentrate on the analysis of maintenance reports. However, companies would benefit from more varied forms of information and knowledge than what appears in maintenance reports. To sum up, customer based information could be collected and utilized more efficiently.

Product-service systems (PSS) first surfaced in the mid 1990s and originated in the environmental studies literature [7]. It was realized that the shift of focus to the final customer need, instead of the product that potentially fulfills the need, could greatly improve sustainability [8]. The business literature also began to show some interest in functional business models; companies started to offer integrated solutions to meet final customer needs. By so doing the companies were able to improve their position in the value chain and their innovation potential, create unique customer relationships and enhance the value added of their offerings [7, 9].

PSS has been defined in many ways. Mont [10] defines PSS as "a system of products, services, networks of actors and supporting infrastructure that is developed to be competitive, satisfy customers and be more environmentally sound than traditional business models". Tukker and Tischner [7] argue that the concept of PSS rests on two pillars; the final satisfaction that the customer wants instead of a product that would potentially satisfy customers' needs is the starting point of business development. The final functionality should also be provided with an open-mind without taking the existing structures, routines and the position of the firm for granted.

PSS enables the fulfillment of customer needs in a more intelligent and efficient manner. It requires companies to increase the value to customers at lower production costs, with lower material inputs and reduced emissions. Companies competing with PSS concepts must constantly improve their capabilities in order to respond to the rapid changes in environment. As a result of these improvements, firms create difficult-to-imitate competitive advantage. The strong relation to customer needs provides companies with chances to re-think their business. PSS concepts enable more radical forms of innovation [11].

Due to the complexity of the PSS concepts for the manufacturer side, the need to design PSS from the customer's perspective and involve the customer in the development process early on [12], companies need customer based knowledge (e.g., customer needs) and new ideas for the development processes to enable the success of the PSS concept. Latent customer needs and radical new ideas are especially important for companies.

In this paper, we propose a framework capable of capturing the knowledge and ideas of customer contacts better than the conventionally used tools and frameworks. The objective of the framework is to capture customer based knowledge and especially the possible radical improvements, better. It is meant especially for employees who have customer contacts. The customer observation framework presented in this study should neither disturb the main duties of the observing employees nor the customer's representatives. The emphasis of this framework is on business-to-business companies in the manufacturing industry. It is, however, general and can be applied to a variety of different business areas.

35.2 Front End of Innovation

The Front End of Innovation (FEI) differs considerably from the later stages of product and service development processes. The main characteristics of an FEI project are uncertainty and creativity, whereas New Product Development (NPD) is more goal-oriented, planned and predictable. The FEI in general consumes fewer resources than NPD and an FEI project is considerably easier and economical to terminate than an NPD project [13, 14].

A model of the front of innovation is presented in Fig. 35.1. The model consists of four tasks: opportunity identification, idea generation, enrichment of the ideas and idea evaluation. The front end is influenced by innovation management, corporate strategy and culture [15]. After the front end, the idea goes to conceptualization and further development. In this paper, the focus is on the very first activities of the front end of innovation, especially on opportunity identification, idea generation and the input knowledge to the process.

The input of knowledge to idea generation and opportunity identification may come both from inside and outside the company. The internal ideas may come, for instance, from the R & D department, sales, marketing and production. The possible media of internal knowledge transfer include idea forum in intranet, enterprise resource planning (ERP), customer relationship management (CRM), inner feedback and discussions. External ideas may come, for example, from customers, competitors, suppliers, consultants and scholarly sources. The possible tools and media include market analyzes, customer satisfaction surveys, tracking of intellectual property rights, scientific publications, feedback and discussions with customers, suppliers and experts.

Good practices in bringing more ideas to the front end of innovation include motivation and social pressure to make ideas explicit. A quickly functioning review system, feedback and progress reports to the person responsible for the idea and demanding that all the ideas should be in use in a categorized form should also

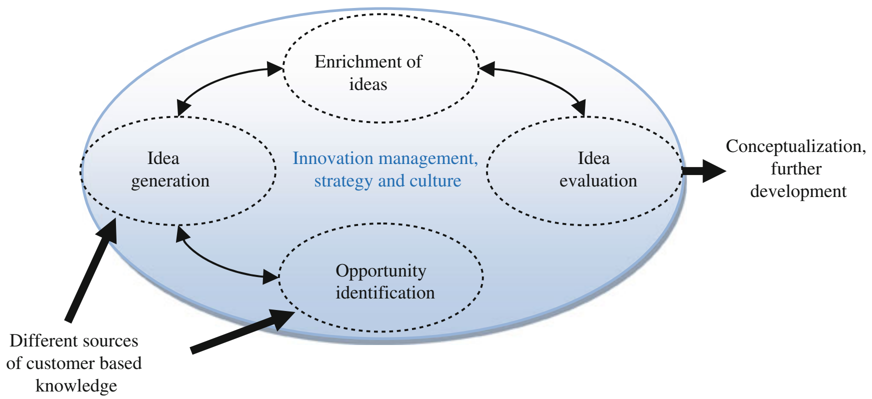


Fig. 35.1 The front end of innovation (adapted from [15])

help in creating new ideas. Other possible good practices are innovation contests in order to direct idea generation to benefit a specific project, theme days for innovation and using ideas from previous projects [15].

Because of the complexity of PSS concepts (integration of product and service components), PSS development models need more feedback loops and communication with different development stakeholders than do conventional product and service development models. The first phase of PSS development (FEI) especially needs more knowledge. The customer's role as a source of knowledge is emphasized in PSS development starting from the front end of innovation. Customer contacts are an important, but underexploited, source of customer based knowledge and ideas.

35.3 Customer Contacts as a Source of Information and Ideas

Customer contacts in this paper are defined as all the contacts the employees of a company have with the representatives and the facilities of the customer. A contact does not have to occur in the same place or even at a same time. It can be achieved through videoconferencing, phone, email or other means of contact.

Frontline employees (employees with regular customer contacts) are seen as an important source of information on latent customer needs, opportunities for innovation [16, 17] and even as the most important source of innovations [2]. Sales and delivery staff have considerable knowledge of customers and the competitive offerings at their disposal [18]. User research is also found to be an effective way of identifying customers' needs [15]. Customer feedback is seen as the activity encouraging service innovations the most [2]. In practice, however, customer interfaces are not utilized as well as they could be. The reason for this deficiency in utilization is partly attributable to the lack of tools for optimally exploiting the interfaces. The frontline personnel may underestimate the importance of the information they could provide and the companies may also view the information, coming from the frontline employees, as unimportant or biased.

Conventional marketing research tools offer little help in anticipating changing customer needs and in creating the breakthrough innovations of the future [6, 19]. In order to better exploit customer contacts, two ways have been proposed in the business literature: active customer participation [20, 21] and observing customers [22, 23]. In this paper, the emphasis is on the latter, i.e., on customer observation.

35.3.1 Customer Observation

The anthropology-based methods for examining the customer in its natural environment have great potential in identifying the latent needs and capturing new innovations [22]. With its anthropological background, customer observation is a method that can be used to support all phases of the innovation process, as well as the delivery and use phases of products and services. However, the emphasis of this study is on the front end of innovation; customer observation as an input that could bring customer-oriented, breakthrough ideas to the innovation process.

Customer observation can be defined as *watching consumers use products or services in the customer's own environment*. [22]. However, the emphasis of this paper is on observing all the contacts with a customer regardless of the medium of contact. It should also be noted that not only the use of products and services is of interest, but also, for instance, the attitudes toward the company or its competitors, the machinery and tools they are using, and investment plans they share, might be of interest. For these reasons, the customer observation in this study is defined as *observing everything of importance related to a contact with a customer*.

35.3.2 Strengths, Risks and Weaknesses of Customer Observation

The customer observation method has some strengths as well as weaknesses and risks in comparison to more conventional marketing research methods. The strengths, weaknesses and risks discussed here originate from a literature review [22–25] and from the implementation phase of the method. Strengths include better chances for identifying innovative and novel ideas. Action is recorded instead of interpreting it. An observer has access to everyday activities and the contact is not limited to "touch points" as in interviews. The understanding of the customer is intimate and not limited by what the customer remembers and is capable of articulating. It is more likely that latent and unmet needs may surface. A competent observer notices deviations from the standard in behavior, organization or systems. The buying behavior, actual product usage, points of customer frustration, and compromises and workarounds developed by the customer may be revealed. Customer is met in its natural environment and is not interrupted from the natural workflow as much as when doing inquiries. Customer observation also promotes learning and sharing of knowledge; all the employees should have access to the observations and the observers should receive feedback on their observations. The increased knowledge about the customer (processes, products, possible problems, etc.) enhances the customer-provider relationship and leads to improved customer orientation.

There are also some risks and weaknesses in this method. The observer may not be able to explain the cause or reason to an action made by a customer. The observation data may be biased and analyzing the data is time consuming. The privacy of the customer should be taken into account as they are not necessarily happy about people making observations on their premises and company. There is a risk that customer observation impinges on the main work of an employee. The implementation of the method may also be difficult as it requires a change in the mindset of most employees.

The abovementioned strengths, risks and weaknesses are taken into account in the creation of the framework for customer observation. In the next paper, the framework is presented and its testing in a pilot project is described.

35.4 A Framework for Customer Observation

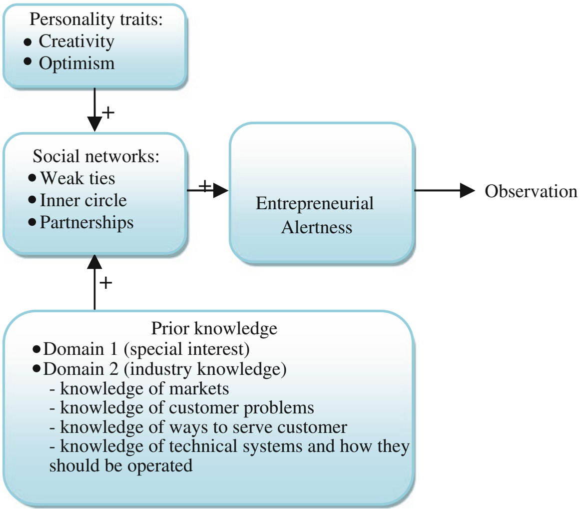
The variants of observation presented in the literature, e.g., [22, 23], treat customer observation as the main responsibility of the observers. However, the framework presented in this study is for all the employees of a company with customer contacts. This means that observing customers is not the main responsibility of the employees, but a back office process which ideally is not apparent to the customer and does not affect the main responsibilities of the employee. The observers (for example service and sales personnel) usually have other priorities, such as the maintenance of a machine or selling a product or service, than just observing their customers. These realities should not diminish the role of customer observation in comparison to more common types of observation. Instead of concentrating on customers' needs irregularly, customer observation is practiced at every customer contact by all employees. In this way, a large number of observations can be gathered from a variety of people observing an even larger number of customers.

35.4.1 Prerequisites of Customer Observation Framework

The guidelines for the customer observation framework are drawn from the entrepreneurial opportunity recognition literature. Entrepreneurial opportunity recognition captures the main aspects of making an observation. It explains the prerequisites necessary for employees to observe or to find an opportunity. The prerequisites of customer observation are presented in Fig. 35.2. The factors influencing the process of observation are alertness, prior knowledge, social networks and Social networks and personality traits including optimism and creativity [26].

Social networks include weak ties (casual acquaintances), partnerships (existing team members) and inner circle (all those with whom the entrepreneur/observer has a long term relationship) [26]. Inner circle and partnerships may have some value as a source of observations, but the most important of social networks are the weak ties. These weak ties can be considered potential sources of observations in customer contacts as they are not as familiar to the observer as stronger ties like partnerships and inner circle: weak ties or acquaintances are more likely to yield unique information than are close friends [27].

Fig. 35.2 The prerequisites for customer observation (adapted from [25])



The term "(entrepreneurial) alertness" is an attempt to explain the observation; the observation is preceded by a heightened state of alertness to information. This heightened state of alertness can be defined as an inclination to be sensitive to information about occasions, objects and patterns of behavior in an environment. The observer is then especially sensitive to maker and user problems, interests and needs still unmet, and new combinations of resources. The higher the level of alertness, the more likely is the opportunity to be recognized. The threshold is likely reached when several factors, such as relevant prior knowledge and experience, social networks and specific personality traits, coincide [26, 28].

35.4.2 The Framework

The customer observation framework is presented in Fig. 35.3. The phases of the customer observation method are requirements elicitation, observation, reporting the observation, analyzing the observation and further development. The goal of this method is basically to streamline the management systems of the company in question so that it is as easy as possible for the observers to reach the threshold of alertness and make observations. After the observation, the goal is to make the observation visible to other employees of the company so that the observation can be distributed to the people who would benefit from it. Through development processes, the observations are then transformed into valuable knowledge, and new products and services.

The main requirements of the customer observation method are employees who are optimistic, creative, have the required prior knowledge and the necessary contacts. These are the factors which should be enhanced to gain the potential

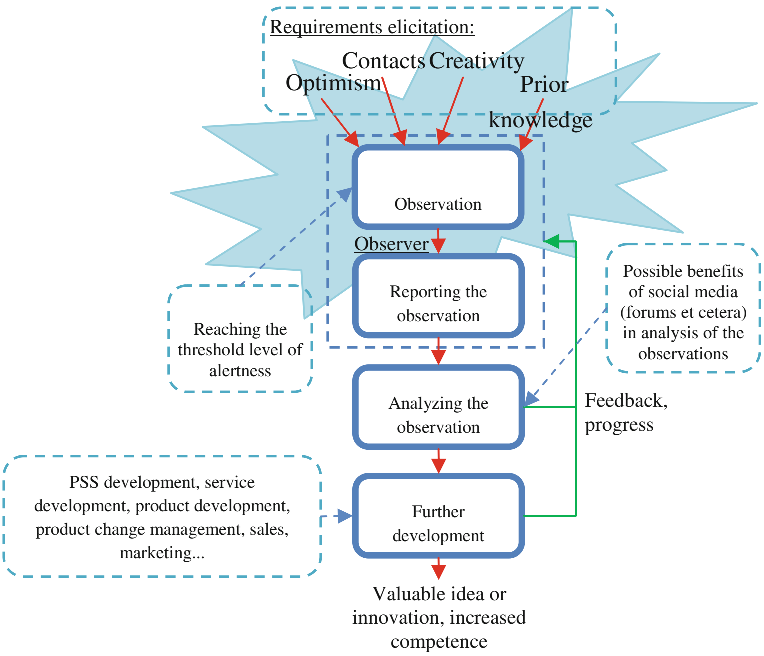


Fig. 35.3 The customer observation framework

benefits of customer observation method. Optimism can be supported by an appropriate organizational culture and means of motivation. Feedback is seen as the single most important factor regarding optimism. Creativity is to some extent inborn, but can also be enabled by communication, and an organizational culture which is open to new ideas and does not penalize possible failures too harshly. Prior knowledge can be divided into personal interest and profession-related knowledge. Both of these categories are important as the points of personal interest bring unique perspectives to the matter in question and may be a source for radical new ideas. However, without adequate knowledge of the technologies involved and the industry in question, the employee may not have enough touch points to make the observation. Prior knowledge can be increased by training, and employing experienced and well-trained staff. Social networks can, to an extent, be enhanced by increased customer contacts (being close to the customer). Of social networks, the weak ties are the most potential source of observations.

When an observer's alertness reaches the threshold level, an observation is made. The level of alertness needed for an employee to submit an observation can be lowered by making the submission of observations as easy as possible. If the observer is motivated and has means of sharing the observation, the observation is made usable for other employees and is transformed into an idea. The observation is submitted to the information system of the company (ERP, new information systems should be avoided for acceptance and complexity reasons), where it is analyzed by another employee and sent on to a person or team that the observation may concern. This stage is crucial for the success of the method as many observations are made but are not necessarily submitted to the system. A good practice would be to submit the observation to a common forum, or to another form of social media, where it can be discussed and developed into an idea. At this point, feedback is given to the observer and the observer can track the progress of the observation from the information system used. For motivational reasons, the feedback should be personal instead of just an automated reply.

The handling of the observation and the development processes should be based primarily on the existing feedback and product development processes for cost and acceptance reasons. The observation may also end up contributing to other departments such as sales, marketing or logistics. The desired outcomes of the method are valuable ideas or innovations, but also increased competences of and communication between employees.

35.4.3 Testing the Framework

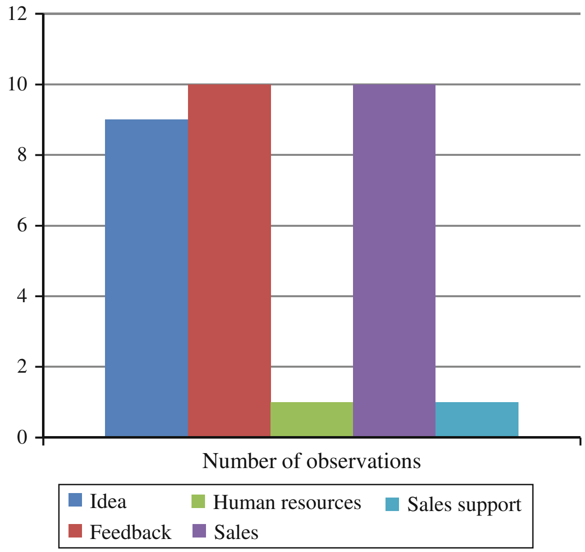
The validity of the customer observation framework was tested on a pilot project in the Finnish service oriented manufacturing company Fastems. The piloting was executed in service departments of the company located in Scandinavia, Germany and Italy in March–April 2011. The participants were trained to the observation method both before and during the piloting.

The goal of the piloting was to find out if the proposed framework is suitable for capturing customer based knowledge, discovering latent customer needs and bringing radical ideas to the innovation process. The employees made observations and submitted them to the information system of the case company, where they were analyzed. The observation functionality was integrated into the existing information system and could be used by all the employees. Prompt feedback was given to the observers about the progress in handling their observations.

A total of 35 observations were received from the observers. If a company with 100 employees with regular customer contacts would make observations for 1 year at this rate, it would result in more than 1,000 observations a year. However, it is doubtful if the employees can submit the observations at the same rate throughout the year. It is certain that the employees make valuable observations all the time, but sharing these may be seen as extra work that is not considered important. Thus, attention must be paid to motivating the employees to share their observations.

The analysis of the observations is very challenging as, for example, an observation classified as a product improvement may ultimately benefit the creation of a new service. The classification was made with the assistance of the company's employees. The majority of the observations were classified as product related, organization related or sales opportunities. Some new product and service ideas were also received. After the observations were classified, they were forwarded to the organizations that would benefit from them. In Fig. 35.4, the

Fig. 35.4 The number of observations by further development process



observations are divided according to their further development process. Because of the similarity of several observations, not all individual observations are sent for further development.

The majority of the observations were forwarded to the feedback process, to the front end of innovation or to the sales department. The difference between feedback and idea processes is that the feedback observations are seen as minor improvements which do not require the idea evaluation process. A substantial share (ca 25 %) of observations forwarded to the front end of innovation is a positive signal about the usefulness of the customer observation framework as a source of ideas. These ideas are based on perceived customer needs and on problems that the customers face, hence increasing the customer orientation of the ideas. The variety in the types of observations is also a good sign as it would suggest that the observation framework yields rich and diversified information.

The results of the piloting show that the customer observation framework is a valid tool for gathering customer based information and new ideas for the innovation process. It seems that the framework can provide latent customer needs and possibly also new radical ideas. However, because of the limited time available for the piloting, convincing evidence cannot be provided.

35.5 Conclusions

Customer co-creation and customer observation are efficient methods for increasing the amount of customer based information in companies. In this paper, we concentrate on the latter case. The customer observation framework presented provides diversified and customer-oriented knowledge. Through increased knowledge about the customer, the framework enhances the customer relationships and promotes the development of successful product-service systems.

Customer contacts may be better utilized by implementing the customer observation framework. The framework has a potential to identify latent customer needs, which are an important source of competitive advantage for product, service and product-service system development. It can provide a large number of customer-oriented observations to serve as a basis for new radical ideas. However, because of the limited time reserved for the piloting the magnitude of the benefits is still uncertain, although the preliminary results were promising. More research needs to be done in order to truly validate this framework. The indirect benefits (e.g., increased competencies and communication) could likewise not so far be verified. An important factor to be taken into account is the maintenance of customer observation after the piloting stage. Attention must be paid to the motivation of the observers, especially by means of providing the observers with sufficient feedback and recognition, and explaining the importance of the customer observation.

Even though customer observation is a very usable framework for identifying customer needs and bringing new ideas into the innovation process, companies also need other sources of information. The customer observation framework works well as an addition to the companies' existing idea and knowledge capturing processes.

Acknowledgments The authors gratefully acknowledge the Finnish Funding Agency for Technology and Innovation (TEKES) for the financial support that made this study possible. We also wish to thank the interviewed companies and especially the case company Fastems.

References

1. Chesbrough HW (2003) Open innovation: the new imperative for creating and profiting from technology. Harv Business Press, Boston
2. Panesar S, Markeset T (2008) Development of a framework for industrial service innovation management and coordination. J Qual Maintenance Eng 14(2):177–193
3. Ahonen T, Reunanen M, Kunttu S, Hanski J, Välisalo T (2011) Customer needs and knowledge in product-service systems. In: Proceedings of the 24th international congress on condition monitoring and diagnostics engineering management, Stavanger, Norway
4. Olsen NV, Sallis J (2006) Market scanning for new service development. Eur J Mark 40(5/6):466–484
5. Nambisan S (2002) Designing virtual environments for new product development: toward a theory. Acad Manag Rev 27(3):392–413
6. Slater S (2001) Market orientation at the beginning of a new millennium. Managing Serv Qual 11(4):230–232
7. 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
8. Schmidt-Bleek F (1993) Wieviel Umwelt braucht der Mensch? MIPS—das Mass für ökologisches Wirtschaften. In 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
9. Wise R, Baumgartner P (1999) Go downstream—the new profit imperative in manufacturing. Harv Bus Rev 77:133–141
10. Mont O (2001) Introducing and developing a product-service system (PSS) in Sweden. IIIEE Lund University, Lund
11. Van Halen C, Vezzoli C, Wimmer R (2005) Methodology for product service system innovation: how to develop clean, clever and competitive strategies in companies. Uitgeverij Van Gorcum, Assen
12. Baines TS, Lightfoot HW, Evans S, Neely A, Greenough R, Peppard J, Roy R, Shehab E, Braganza A, Tiwari A, Alcock JR, Angus JP, Bastl M, Cousens A, Irving P, Johnson M, Kingston J, Lockett H, Martinez V, Michele P, Tranfield D, Walton IM, Wilson H (2007) State-of-the-art in product-service systems. IMechE Part B J Eng Manuf 221(B10):1543–1552
13. Koen P, Ajamian G, Boyce S, Clamen A, Fisher E, Fountoulakis S, Johnson A, Pushpinder P, Seibert R (2002) Fuzzy front end: effective methods, tools and techniques. In: Belliveau P, Griffin A, Somermeyer S (2005) The PDMA toolbook for new product development. Wiley, New York
14. Jongbae K, Wilemon D (2002) Focusing the fuzzy front-end in new product development. R & D Management 4:269–279
15. Apilo T, Taskinen T (2006) Innovaatioiden johtaminen. Espoo, Finland: Research Notes 2330. In Finnish
16. de Jong J, Vermeulen P (2003) Organizing successful new service development: a literature review. Manag Decis 41(9):844–858
17. de Brentani U (2001) Innovative versus incremental new business service: different keys for achieving success. J Prod Innovation Manage 18:169–187
18. Gebauer H, Krempl R, Fleisch E (2008) Service development in traditional product manufacturing companies. Eur J Innovation Manage 11(2):219–240
19. Matthing J, Sandén B, Edvardsson B (2004) New service development: learning from and with customers. Int J Serv Ind Manage 15(5):479–498
20. Prahalad CK, Ramaswamy V (2004) Co-creating unique value with customers. Strategy Leadersh 32(3):4–9
21. Sawhney M, Prandelli E (2000) Communities of creation: managing distributed innovation in turbulent markets. Calif Manage Rev 42(4):24–54
22. Leonard D, Rayport J (1997) Spark innovation through empathic design. Harv Bus Rev 75(6):102–113
23. Rosenthal SR, Capper M (2006) Ethnographies in the front end: designing for enhanced customer experiences. J Prod Innovation Manage 23:215–237
24. Wägar K (2008) Exploring tools for learning about customers in a service setting. Int J Serv Ind Manage 19(5):596–620
25. Krueckeberg H (1989) Supermarket customer observation and electronic data analysis with implications for the marketing plan. J Food Distrib Res 20(1):74–90 http://econpapers.repec.org/article/agsjlofdr/27913.htm
26. Ardichvili A, Cardozo R, Sourav R (2003) A theory of entrepreneurial opportunity identification and development. J Bus Ventur 18:105–123
27. Granovetter M (1973) The strength of weak ties. Am J Sociol 7(6):1360–1380
28. Ray S, Cardozo R (1996) Sensitivity and creativity in entrepreneurial opportunity recognition: a framework for empirical investigation. Presented at the sixth global entrepreneurship research conference, Imperial College, London. In: Ardichvili A, Cardozo R, Sourav R (2003) A theory of entrepreneurial opportunity identification and development. J Bus Ventur 18:105–123

Chapter 36
Application of a Unified Reference Model Across Asset Types: Comparative Cases

Y. Wijnia, J. de Croon and J. P. Liyanage

Abstract Asset Management has grown over the last few years and has begun to use many terminologies and concepts across different applications. This has had notable negative effects on the further growth of the discipline, as the present conditions only contribute to complexity and chaos rather than leading the way toward a commonly acceptable approach. In this context, a common reference model becomes extremely useful that can provide the necessary elements which can be capitalized for learning and development efforts for wider applications. The issue here is to get the principles intact that has comparable and defining features across many contexts. In this respect some initial efforts were invested to develop an approach for a unified asset management reference model. This paper brings further work on this reference model into perspective by discussing the application of the reference model across different asset types. The purpose here is to communicate the potential of such a unified approach as a valuable foundation to build on.

Y. Wijnia (✉) · J. de Croon
Asset Resolutions B.V, 30113, 8003 CC Zwolle, The Netherlands
e-mail: ype.wijnia@assetresolutions.nl

J. de Croon
e-mail: john.de.croon@assetresolutions.nl

Y. Wijnia
Faculty of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands

J. P. Liyanage
Faculty of Science and Technology, University of Stavanger, Stavanger, Norway
e-mail: j.p.liyanage@uis.no

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_36,

36.1 Introduction

Managing assets in the modern industrial climate has begun to provide different realms for different asset managers in different sectors. The traditional thinking and the use of varied approaches have only contributed to creating disparities rather than assimilating the development solutions. This to a certain extent is not a matter of the complexity of the subject but rather the immaturity behind the thinking process on what asset management really is about, and the lack of insight to the deeper meanings of the roles of assets and the functions of asset managers in different contexts. Concepts and terminologies that are largely seen used today have failed to provide the level of maturity required. In many industrial minds, thus unfortunately, asset management remains a hypothetical and an academic solution rather than embracing it as a timely relevant and value creating concept.

To shed some light in this regard, some initial work done was discussed by Wijnia et al. [1] where attempts were made to develop the basis for a unified reference model. This is based on the understanding that despite the obvious differences between the fields, there seems to be a deeper unexplored commonality. Given that every asset manager is in his or her own way and context trying to optimize value, many application fields have the potential to derive effective solutions from a single generic concept. In such a case, lessons may well be learned between know-how across the fields as well as sectors. Such pragmatically usable generic concept should not be too far out from the realm in sectors. If the cash flow is the only idea shared between the fields, there is not much to learn. In this paper, the idea of such a generic framework is furthered by structuring the basic framework developed across different asset types. The purpose of this second phase of the study was to explore the potential of the initially developed reference model as a common ground.

36.2 Structuring Types of Asset Management by the Framework

For this particular task three types of assets were considered, namely infrastructure, production facility and financial. Those assets have specific attributes of their own and at the same time follow management processes that are regulated by relatively different contexts. This provided a fruitful basis to have a meaningful insight into the realm of asset types using the common reference frame.

36.2.1 Infrastructure Asset Management

Infrastructures are, in the most conventional definition, technical systems that provide paths between points. The plural is on purpose. A technical system providing a path between two specific points is “just” a link, a connection between

A and B. The interesting thing about infrastructures is that they connect A not only to B, but also to C and so on. Typical examples of such conventional infrastructures are roads, railways, waterways, the drinking water system, sewage systems, electricity and gas distribution grids and telecom grids. Despite technical differences, infrastructure assets seem to share many characteristics:

1. They have a long to very long lifespan (sometimes more than 50 years)
2. They are costly but cannot be sold for another purpose (how to sell an overpass)
3. They mostly are passive elements, they just provide a path; only the controls (switches and so on) are operated
4. Because of longevity not all records of current assets survived: the asset base is not fully known
5. Infrastructures are evolutionary systems, growing over time with sequential add-ons
6. They are networked, with little or no key elements that can bring down the system
7. They are dispersed and operated in the public domain
8. The owners/decision makers are not the users
9. Infrastructure users can be highly anonymous, opening the door for misuse of the assets.

Based on discussions with Dutch infrastructure asset managers [2, 3], no specific key lifecycle phase for all infrastructure assets could be indicated. For some assets the focus is on design in terms of capacity and route (cables, pipelines, roads), for others there is more attention for maintenance (switches, road surface) or operation (switches). In general, passive elements are mostly design and perhaps a bit maintenance, for active elements the equipment is generally of the shelf, with more attention for maintenance and operation.

The environment for most infrastructure assets tends to be very stable. The infrastructures are the facilitators of some very basic needs (power, drinking water, mobility), and their volume does not show much volatility, but is highly correlated to the population size and state of the economy. An exception can be made for the telecom companies, which have seen major changes in the last decades (mobile phones, internet). However, even though the growth was significantly larger than in other infrastructures, the growth has been consistent. The market prices are also stable compared to commercial industries. In many cases, infrastructures are monopolies and the income is determined by a regulator or some other governmental agency. Again, this does not apply to telecom, as there is competition between infrastructures in that domain, with accompanying pressure on prices. For most infrastructures is the development of technology rather slow. Cables, pipelines and roads are old technologies. There is continuous improvement in cost per unit, reliability and so on, but breakthroughs are rare. Not surprisingly, telecom is the outlier again. A major technology change has passed (copper to glass), and the services provided over the system are not comparable with the old.

For most infrastructures, management attention not only focuses on the financials. Reliability and safety matter as well. The reason is quite straightforward: as the assets are in the public domain, any incident will be visible. Ill-performance with regard to safety and reliability will draw attention and may result in undesired involvement of government officials, or even the shareholders, who tend to be government as well. Over the years the operators learned to deal with it by internalizing the public values. This is especially true for electricity, gas and water. The reliability of those systems is extremely high (over 99.9 % [4, 5]), potentially beyond the point of the economic optimum. With regard to the safety, despite the potential of respectively electrocution, explosions and poisoning, a very high performance level has been reached. In Fig. 36.1 important characteristics of infrastructure asset management are presented.

36.2.2 Production Facility Asset Management

Production facilities are in essence locations where resources are converted into products. The term product is used in a broad sense, including both goods and services. In this view, a call center (which converts knowledge and human resources into answers, advice or complaint resolution) is a production facility as well as steel mills (which convert raw materials like coal and ore into steel). Other examples are oil and gas refineries, chemical plants, discrete manufacturing (cars, electronics), food processing, electricity generators and defense industries.

A binding characteristic of production facility assets is that they require operation: they are active assets, not passive assets. If any of the needed resources (power, feedstock, operators) lacks, nothing comes out. All production facilities for goods use rotating equipment. Rotating equipment is vulnerable to wear and tear. Depending on the task and operating conditions, rotating equipment may function

Asset characteristics • Function: provide path for the public • Behavior: mostly passive, they do not require operation to function, nodes may be active • Life cycle length: long to very long for equipment, the geographical routes tend to be unlimited • Location: Fixed, distributed, line assets (links), linear assets and point assets (links, switches), in the public domain • Type: Electrical, civil, mechanical	**Environment** • Market volume: very stable • Market prices: Very stable • Technology development: slow
Key life cycle phase • Key life cycle phase: o Passive elements: design o Active elements: operation	**Management attention** • Main activity: investment planning for passive assets, maintenance for active assets • Analysis tools: demand prognosis • Value system: reliability, safety, as long as affordable • Asset control: partial, asset users are not the owners or managers

Fig. 36.1 Characteristics of infrastructure asset management

for hours or days (drill head in mining) up to years (pumps, ventilators, transportation belts). Non rotating equipment (reactor vessels, ovens, frames) typically lasts much longer. Another typical characteristic of production facilities is that they are confined physical locations, at least for goods. Services may be produced dispersed as well (facilitated by modern communication technology, which means a call center operator may work from home), though the investments needed to do so and the lack of direct control an incentive to work on a confined facility. Production facilities are often designed at the start, though their design may change significantly over the years. Furthermore, production facilities may be sold, both in terms of superfluous equipment (after an upgrade for example) or as a whole. The latter will require some due diligence, for a second hand equipment markets exist.

For production facilities, the operational phase is everything. The design needs to be right, but in general any production facility will see significant upgrades in capacity during its lifetime. It is not unusual for chemical plants to outcompete their design capacity by multiple times, by optimization of the operational parameters. And in discrete manufacturing processes, the continuous improvement of the operation may result in a significant decrease of the takt time. The Scania truck manufacturing plant in Zwolle for example reported a 40 % increase in the productivity of the location in 18 months by optimization of the production process [6].

The environment for production facilities can be range from stable (though not as stable as for infrastructures) to very dynamic, which strongly depends on the product. Basic needs like energy and food will have a more stable environment than consumer goods like electronics. Interestingly, the lifecycle of the production assets may be much longer than the product lifecycle.[1] The production facility for cars for example can produce many different models over the years with the same equipment. On the other hand, in computer chip manufacturing, the capacity of the chips strongly depends on the resolution with which the circuits can be printed. Every new generation of chips therefore requires a new production asset. Remembering Moore's Law of capacities doubling every 2 years or so equipment can become obsolete very fast.

As mentioned, the key phase is the operational phase, so that is where management attention is. But it is as much on increasing performance with current assets as it is on reducing the risk of malfunctioning assets, the role of classical maintenance. But because of the omnipresence of rotating equipment, maintenance cost are significantly higher than those of infrastructure assets, usually in the order of 1–2 % of asset value [7]. The assets exist to make products and produce income, therefore the most important value is the financial one. Reliability can (should) be judged economically and is thus financial. Safety, compliance, sustainability and the like are treated more like constraints than as true values. After all, the facility is

[1] The product lifecycle is the time the product is available on the market, not the life expectancy of the product itself. A specific computer model may only be sold for a few months, whereas the computer itself may function for years, and the production facility for tens of years. On the other hand, disposable cups have a useful life expectancy of a few minutes, whereas the product may be on the market for years.

Asset characteristics	Environment
• Function: Produce goods for the owner to sell • Behavior: active, assets require input to function • Life cycle length: short/medium (discrete manufacturing) to long (chemical industry) • Location: concentrated, fenced off • Type: Electrical, civil, mechanical	• Market volume: reasonably stable (food, oil & gas) to volatile (electronics). • Market prices: stable (food), to volatile (oil and gas) • Technology development: fast (e.g. mobile phones) to slow (oil and gas)
Key life cycle phase	**Management**
• Key life cycle phase: o operation	• Main activity: maintenance management • Analysis tool: past performance • Value system: financial: reliability, safety are constraints time to market? • Asset control: full, owners are the users though activities can be outsourced, assets are fenced off

Fig. 36.2 Characteristics of production facility asset management

fenced off, and most incidents stay within the fence. Yet, for production facilities (or rather the products they make) image can be very important. Shell lost a significant share of market volume in 1993 when they decided to sink off the obsolete oil rig Brent Spar. Nevertheless, in most cases the asset management decisions hardly influence the image of the product. In Fig. 36.2 characteristics for production facilities are summarized.

36.2.3 Financial Asset Management

Financial asset management deals with tradable securities. These can be any form, ranging from common stock, bonds and commodities to more complex derivatives like swaps, options and futures. Securities can be direct like a share of a company or a treasury bond, but also indirect, like an investment in an index fund. The liquid markets provide every opportunity to construct a portfolio of assets that matches the risk profile of the investor, though it can be inconvenient for smaller investors to do so. Therefore, there is a large volume of pre-packaged assets like the already mentioned index funds. In some cases the assets are securitized debts, like collateralized mortgage obligations. Financial asset managers generally manage a portfolio of securities (to keep in line with the risk profile of the investor), but there is no attachment to the securities in the portfolio.

Financial assets are characterized by a number of things. Most importantly, the assets are traded on very liquid markets.[2] This removes the need for managing the underlying securities, as they can be sold immediately if underperforming.

[2] The average daily volume on the New York stock exchange was 153 billion dollar, about 1% of the annual GDP of the United States of America.

Furthermore, the assets are highly dispersed. A portfolio can consist of shares in companies all over the world, and because of modern communication technology, asset managers can get involved in all markets of the world. Thirdly, the value of the assets is determined by anybody trading anywhere in the world, as the asset price is the average price of all transactions in the last timeframe. Normally, this is just a bit of noise, but if everybody acts coordinated the effects can be dramatic (like the credit crunch of 2007–2010).

Within financial asset management, it is difficult to pinpoint the lifecycle phase as it strongly depends on the viewpoint one has. For example, if the portfolio of securities is regarded as the asset, it seems fair to state that the key phase is operation and maintenance. The performance of the parts of the portfolio is monitored, and underperforming parts are replaced. Given the enormous value of the trade compared to the value of the underlying assets, it seems equally fair to state that the design and construction of the portfolio is most important, as the securities are not replaced by the same securities, but by something completely different. The portfolio is so to say continuously redesigned. In the context of this paper the maintenance perspective is used.

The environment can be extremely volatile. This is not only true for individual securities, but for the portfolio as well. Market averages can drop over 20 % in a single day of trading.[3] Volatility may have been increased by modern communication technology, to that asset managers can react faster to movements in the market. However, the production technology did not change very fast. The basics of a portfolio are still shares and bonds, though they may have been repackaged into exotic assets.

Management attention in financial asset management is on buying and selling assets. Because of the limited share the asset owner generally has in a company, it is difficult to control the management of the underlying asset directly. And there is no need to either. Because of the high variety of securities traded today it is much faster to change the portfolio.

Financial asset management is purely about financials, both in terms of profit as in terms of risk. Any other value is indirect at the most. Some funds may focus on sustainable investments, but if they do not deliver financial value they will not generate profit for the investor. However, there is a significant amount of emotion into the asset value. If traders “believe” an asset will do well in future, it will do well tomorrow as the high demand for shares will drive up the price. This may result in stock market bubbles where securities are traded well above their fair value, waiting for the inevitable collapse. A famous Dutch example was the Tulip Mania of the 1630s [8], though there is debate on the truth of this. Other examples are the Mississippi Scheme or the South Sea Bubble. This emotional response opens the door for malignant speculators to manipulate the value of their securities by providing false information. In Fig. 36.3 characteristics for financial asset management is presented.

[3] On the 19 October 1987 the Dow Jones index dropped 22.6 % in a single day.

Asset characteristics	**Environment**
• Function: produce income for the owner • Behavior: active • Life cycle length: extremely short: the portfolio can be reconstructed overnight • Location: distributed, not fenced off. Any trader around the world • Type: Financial, packaged rights to earnings of some underlying real assets	• Market volume: reasonably stable (food, oil & gas) to volatile (electronics) • Market prices: stable (food), cyclic (e.g. iron, copper) • Technology development: fast (e.g. mobile phones) to slow (oil and gas) is applicable for applied infrastructure, not for financial assets itself
Key life cycle phase	**Management**

• Key life cycle phase:
  o Design: packaging securities to form a new type of asset
  o Operation: managing a portfolio of assets
	• Actions: portfolio composition
	• Analysis tool: historic prices though in some cases the underlying asset is analyzed
	• Value system: financial
	• Asset control: partial, portfolio can be changed, but the value of the underlying assets is subject to actions of all participants in asset trade

Fig. 36.3 Characteristics of financial asset management

36.3 Differences and Similarities

If the characteristics of these different asset systems are compared, differences and similarities can be noted. The interesting thing is that there is no clear order. Production plant assets are in some aspects quite similar to infrastructure assets: they are big, lumpy technical things. But with regard to the value system plant asset management is more like financial asset management: they are both in the private domain to generate income for the owner. This would suggest that on a continuous scale plant asset management is somewhere between infrastructure and financial asset management. However, in some respects infrastructure assets are more like financial assets. They both are highly dispersed and open for external influences. Plants assets are fenced off from the outside. This would suggest that infrastructure asset management is between plant and financial asset management, which is contradicting the previous ordering. The only viable solution is to say that the three forms of asset management are the angles of a triangle, as shown in Fig. 36.4.

Parallel to the sides, there are the demarcation lines: infrastructure assets are public, whereas securities and plants are private in Western countries. Securities on the other hand are liquid assets, whereas plants and infrastructures are fixed. Plant assets are also special, as they are confined, whereas securities and infrastructures are dispersed.

Key question now is whether the scales are dichotomy or continuous. Are there assets that are half way the sides? To start with the liquid to fixed scale, examples are real estate and rolling stock. They are real, technical assets, but on the other

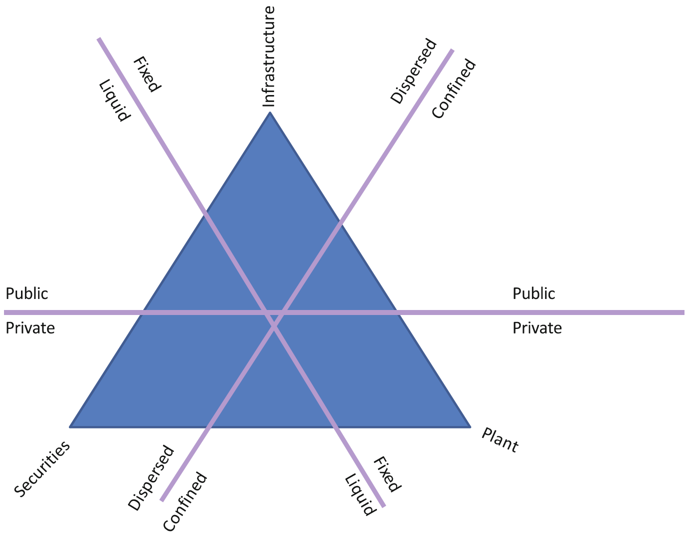


Fig. 36.4 Mapping extremes in asset types along with their distinguishing factors

side markets exist in which they can be traded. However, the time needed to sell a house or a car is most likely much longer than that of selling a financial asset, which is basically a mouse click. This may suggest that the liquidity of the market is the true distinguishing factor, but that in itself depends on the uniqueness of individual assets. Markets in which highly standardized asset are traded are much more liquid than the markets for unique assets as the buyer of a standardized asset knows exactly what the asset is, and for unique assets some more research is needed. But as a proxy, liquidity may do very well.

The second scale is that of dispersed versus confined assets. One example of something in between is a collection of plants as owned by a multinational. Each plant is confined, but the plants are scattered around the world. And due to political instability, even fenced off plants may be very vulnerable in some countries. Another example is a fenced of private infrastructure, like the "leidingstraat" in the Netherlands, which is a fenced of zone for large pipelines.

Finally, the boundary between public and private is crossed by for example the public private partnerships like DBFM contracts in road infrastructures. A private party then designs, builds, finances and maintains the road, and gets paid by the public authority according to how well the road performs. For the private partner it is an asset to generate income but for the authority it delivers public value, if the right key performance indicators are set.

However, it seems difficult to think of any asset truly in the middle, with characteristics somewhere between fixed an liquid, public and private, and dispersed and confined. If there is no such asset, it is not surprising there is no common framework for asset management, as it has no natural anchor point. And any deviation of the center will result in asset management addressing issues that simply do not match characteristics of all assets around the world. Think about selling infrastructures, lubing securities or voting about the production scheme of a chemical plant: it just does not work.

36.4 Conclusion

In this paper, the total realm of asset management has been explored. The forms discussed were infrastructure asset management, production facility asset management and financial asset management. An interesting finding was that these types of asset management have no natural order: they can be both an extreme as be in the middle between the other two. Financial asset management is like infrastructure asset management and unlike production facility asset management in the lack of confinement of the assets, but it is the other way around in terms of the value system. Infrastructures and production facilities share the lumpiness of their assets, contradictory to the liquid assets in finance. Yet, a true unifying asset sharing characteristics with all three forms is difficult to envision. Therefore, in terms of concrete management actions similarities seem to be missing. Further research has to be done to arrive at a true common framework beyond the abstract process model.

Acknowledgments Many insights were developed in the meetings with the Eurenseam Network in Sevilla (2009) and Delft (2011). Ideas have been tested in an interview round conducted in the Netherlands in 2010. The authors wish to thank EURENSEAM members and the interviewed asset managers for their contribution.

References

1. Wijnia YC, de Croon JAW, Liyanage JP (2011) Towards an asset management reference model: basis for a unified approach. In: Proceedings of the 6th world conference on engineering asset management (WCEAM), Cincinnati, 2–4 Oct 2011
2. Van der Lei T, Wijnia YC, Herder PM (2010) Towards an integral asset management framework using asset characteristics, asset environment and lifecycle phases as leading principles. In: Proceedings of the 5th world conference on engineering asset management (WCEAM), Brisbane, 25–27 Oct 2010
3. Wijnia YC, Herder PM (2009) The state of asset management in the Netherlands. In: Proceedings of the 4th world conference on engineering asset management (WCEAM), Athens, 28–30 Sept 2009
4. Netbeheer Nederland (2010) Betrouwbaarheid van elektrictieitsnetten in Nederland 2009. Arnhem
5. Netbeheer Nederland (2009) Storingsrapportage gasdistributienetten 2008. Arnhem
6. Scania (2008) Scania King of the road. The Hague
7. Balk M, De Croon J (2004) Praktische prestatie-indicatoren voor de onderhoudsfunctie, in Maintenance Totaal. De Boer SL, De Croon J (eds). Kluwer
8. Mackay C, Tobias A (1980) Extraordinary popular delusions and the madness of crowds. Harmony Books, New York (First published in 1841)

Chapter 37
Clouds of Technical Data Promise Aid to Asset Management

P. Johnson

Abstract Cloud storage and computing continues to present itself as a solution to the data storage and computational needs required by Asset Management applications. Indeed, a fleet of assets that are properly monitored using sensory networks will create terabytes and even petabytes of data and information. The largest computer and technology suppliers are investing billions of dollars in the evolution of cloud computing technology. To take advantage of cloud computing for Asset Health Management and prognostics it is necessary to understand cloud technology and its challenges. This paper introduces cloud concepts and relates them to the needs of Asset Management.

37.1 Introduction

Asset Management depends on knowledge of the asset's performance, mechanical health, expected use, and expected failure modes. Each of these topics depends on data and algorithms to transform asset data into information, then into knowledge, and then into an action that balances safety, quality, and financial impact. When considering a fleet of assets, the need quickly arises to have terabytes of storage space for data and many cores of computational power to process the data. While the cost and performance of desktop and engineering workstations continues to improve, the need to use a supercomputer for fleet wide data and performance assessment raises the question of infrastructure costs. More and more information technologists are turning to Cloud computing to alleviate the cost of locally hosting a network of computers to serve the supercomputer needs of asset management for a fleet of machines.

P. Johnson (✉)
National Instruments, 11500 North Mopac Blvd, Austin, TX 78759, USA
e-mail: preston.johnson@ni.com

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_37,

To best understand how the Cloud can be used, it is necessary to understand cloud technology and the tools that exist to take advantage of the technology. The largest computing technology companies in the world are spending billions of dollars on the technology. Microsoft, Dell, AT&T, IBM, Google, Apple, Facebook, Siemens, HP, and so on are spending significant portions of their research and development budget on cloud technologies.

A common definition of the Cloud includes hardware, networks, storage, services, and interfaces which combine to provide aspects of computing as a service. The National Institute of Standards (NIST) defines the cloud as "a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [1]. The Gartner Research team expands on this definition by adding five key attributes of the Cloud. These are on-demand service, scalable and elastic, shared, metered by use, and built on standard internet technologies [2]. Fundamentally, then the Cloud is a network of server computers with storage and virtualization software that is accessible over the Internet. Its behavior ultimately depends on the software that runs on the Cloud.

37.2 Types of Clouds

Building on the core infrastructure, the Cloud is commonly divided into three types: Infrastructure as a service, Platform as a service, and Software as a service, Fig. 37.1. Infrastructure as a service provides the fundamental components of the cloud. The customer of this Cloud type provides its own operating system, database, and application software. The information technology community (IT) tends to favor Infrastructure as a service Cloud types as it alleviates the need to own and maintain computing hardware, yet still requires IT skills and personnel. The Platform as a service and Software as a service Cloud types progressively provide more and more of the IT function. In Fig. 37.1, the functions in the yellow box are self managed, managed by the IT team for the benefit of its customers.

When an organization invests in Infrastructure as a service (IaaS), it is outsourcing the use of computers, storage, memory, and network access. Virtualization software allows the customer (of IaaS) to use a dynamic pool of servers and storage. The computing power, memory, and storage space expand and contract as the customer uses the resources. The customer remains responsible for the operating system and all applications software the customer organization will use. Historically, IaaS has been referred to as "managed hosting". Amazon's web services, Rackspace, GoGrid, and VMWare are common suppliers of IaaS Cloud Services.

When an organization invests in Platform as a service (PaaS), it is purchasing IaaS plus an operating system, data storage software infrastructure or database, and programmable interfaces to the operating system and the data storage mechanisms.

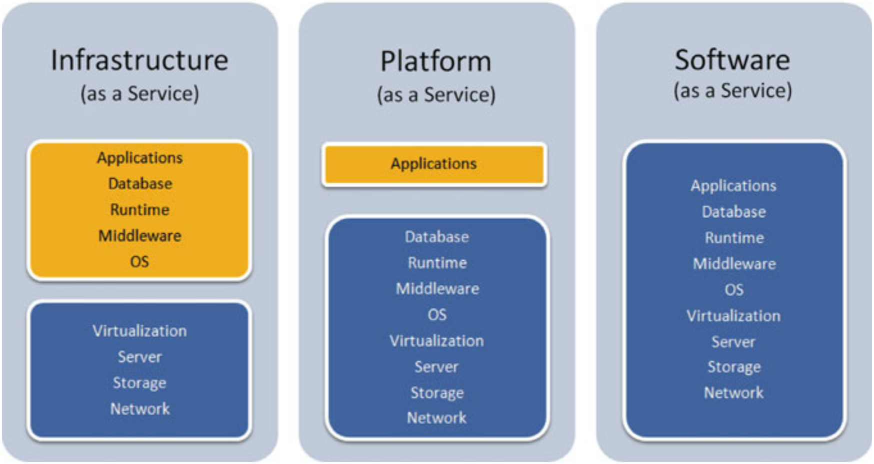


Fig. 37.1 Three types of cloud services

The customer organization then writes an application that works within the provided operating systems and with the provided data base services to manage and process data on behalf of requests from internet clients. The customer is no longer responsible for operating system or data base infrastructure administration tasks. The customer uses the programing language or toolbox provided by the PaaS provider to develop applications for its internal use and/or its customers use. Example PaaS suppliers include Google's App Engine, Microsoft's Azure, and Cloud2Cloud.

When an organization invests in Software as a service (SaaS), it is purchasing PaaS plus pre built applications. SaaS suppliers do offer some customization options to tailor the applications. Historically SaaS services were referred to as Application Service Providers, or ASPs. Example SaaS offerings are Google Applications (Google Documents), salesforce.com, Lithium, TurboTax on-line, Microsoft's Office 365, and Netflix movie streaming services. Essentially with SaaS, the customer accesses an application in the Cloud, inputs or otherwise provides data, and gets reports, analysis, or documents in return.

Choosing the Cloud service for the organization comes down to the intended use. If the organization simply wants a large amount of data storage, perhaps with a simple limited user application, then the IaaS may be the right choice. If the organization wants to develop an application that serves a varying load of end users and wants to take advantage of scalable processing in addition to scalable data storage, the PaaS may be the right choice. If the organization needs a pool of ready to run end user office productivity applications, or work group software, then the range of SaaS options may be the right choice. In choosing, one will find the Cloud to really be a Cloud of Clouds where the organization purchases services from several vendors to provide a mixture of functionality for its needs.

37.3 The Needs of Asset Management

At the core of asset management, is data collected in and around the asset which when interpreted yields information and knowledge that is used to drive action, including maintenance actions. The Cloud then, operating as an asset management tool, must received data, interpret data, and deliver information and knowledge to the owner of the asset. Some of the larger industrial and automotive suppliers do provide a SaaS like asset management service for their customers. Examples include General Motor's On-Star™ service, Boeing's aircraft fleet services GoldCare™, and General Electric's remote monitoring and diagnostic service. These example Cloud services are specifically designed to meet the needs of particular assets and the needs of particular customers.

To serve a broader audience, a configurable asset management framework may be desirable. If an asset management team or Services Company is to construct a configurable Cloud application for use with a variety of assets and end customers, a PaaS provider may be the right choice. Here the asset management design team does not need to worry about the operating system or the data storage, and can focus on the application of interpreting incoming data and creating information and knowledge for end customer use.

To start, it may be convenient to describe a data format which can be supported by data recording devices local to the asset, and that can be transferred to the Cloud application. The data format used should support both summary data, commonly referred to features, as well as original time data, commonly referred to as time waveform records. An example structure is shown in Fig. 37.2.

One advantage of this data format is its ability to be shaped on demand to include addition feature trends, additional waveform records, and even analysis result records. This ability removes the need for an IT resource to continually update database tables, fields, and relations. Each recording from a remote data recording device is self contained in a single data file. An indexing service runs on the PaaS cloud to identify specific data items an asset management application would need to access.

By collecting periodic recordings from individual assets, as well as multiple assets in a customer fleet, or fleets of multiple customers, data becomes

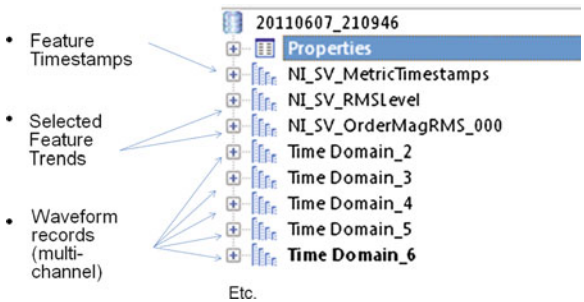


Fig. 37.2 Representative data format for asset management data recording

statistically rich as the volume increases. As with an IaaS, the PaaS data storage resources will expand and contract as more data is added and as older data is removed or archived. Yet, the next component of the Asset Management Cloud application is the analysis software which runs in the Cloud and uses the data collected in the Cloud. In simpler cases, it is possible to simply take an existing single computer application platform and host or execute the application on the Cloud computer resource. This hosted application is able to recognize the arrival of a new data recording, interpret the recording to ascertain information, knowledge, and suggested actions. The only remaining item is communications with the end users of the asset management Cloud application. Communication can be in the form of email alerts with links to on-line reports and graphics, Adobe™ PDF documents created in the PaaS Cloud, or even the download of data files or collections of data files for local off-line analysis.

Communication in itself has several components. These include user management and security. The Asset Management Cloud service needs to provide a form of security of data transfer as well as authentication. Fortunately, these components are often available from the PaaS Cloud provider. For example, HTTPS is a secure form of common internet HTTP traffic. A depiction of an Asset Management Cloud application is shown in Fig. 37.3.

An Asset Management Cloud application can then be created using PaaS Cloud services by carefully choosing a data format, in Cloud application, and leveraging the infrastructure of the PaaS platform service for communications authentication and secure communications.

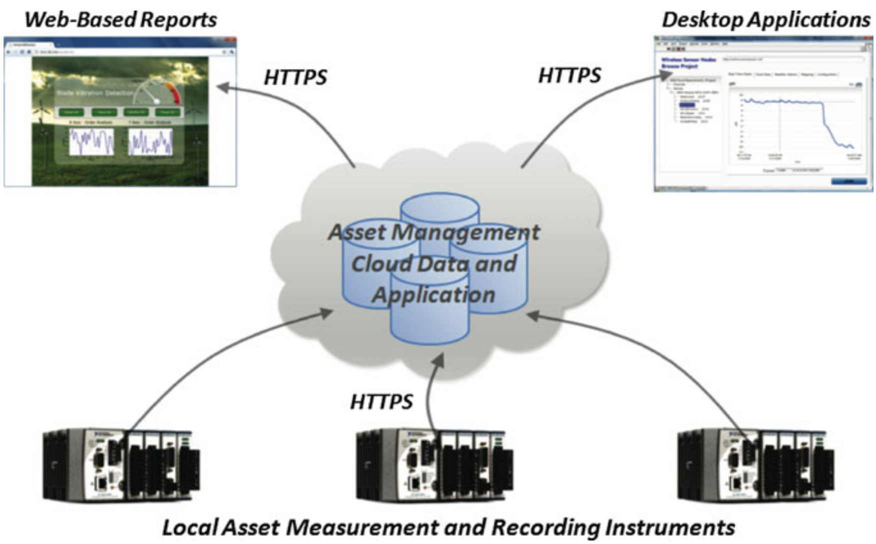


Fig. 37.3 Asset management cloud data and application

United Grinding Technologies (UGT) has created its equipment and production monitoring Cloud application entitled CAREview™ [3]. A graphic of the end user view of the data is shown in Fig. 37.4.

The CAREview™ application receives data from the customers' machines and provides real-time operational views of the data. Machine status, accumulated part production, and even electrical power draws are just some of the information view that is available from most web browser devices. Reports from the "Cloud" application can be customized per end customer requirements.

37.4 Future Needs of Asset Management

However, using an application developed for a single computer does not expand to operate across multiple computers without significant computer science investment. If the Asset Management application becomes popular, more data will be recorded and more complex algorithms will be utilized. Many asset health management engineering professionals already are designing complex matrix math based algorithms that are difficult to execute on a single computer. These algorithms desire access to large amounts of data, which the cloud provides, and extreme computing power. Unfortunately, none of the three Cloud types offer an applications development environment that can scale from one to many computers and back with the ebbs and flows of demand.

As noted during the introduction, key players in Cloud computing technologies are investing billions of dollars in Cloud tools and infrastructure. In the near future, an Asset Management or Prognostics and Health Management (PHM) platform will be desirable by the many equipment services and reliability teams that will both allow customization of asset management software, and retain the elasticity of data storage and computational power that the Cloud provides. If the efforts of the big investors and of those building tools for data recording and asset management in the Cloud, there is likely to be such an application platform. This future platform might build on a PaaS platform like that of Microsoft's Azure™ and take advantage of both data store and computing resources. Perhaps such an application

Fig. 37.4 CAREview™ (Credit UGT and American Machinist publication)

platform would provide an algorithm execution engine that would span across multiple computers as needed by depth of data and complexity of the algorithm. One can only expect these facilities to present themselves.

37.5 Conclusion

The network is full of resources to facilitate the science of Asset Management. The Cloud is the newest of these facilities and promises to provide both data storage and computing power necessary to provide asset management for fleets of machines, vehicles, buildings, etc. Data structures, algorithms, and communications are available now to take advantage of the Cloud. With ongoing investment by the major commercial players, it is likely that a general purpose prognostics application platform will become available which may asset management teams can easily take advantage of.

References

1. GSA, Info.Apps.Gov (2009) What is the Cloud? Available from: http://www.info.apps.gov/content/what-cloud
2. Gartner Research (2009) Gartner newsroom: gartner highlights five attributes of cloud computing. Available from: http://www.gartner.com/it/page.jsp?id=1035013
3. American Machinist (2011) UGT Starts could-based monitoring for grinding. Available from: http://www.americanmachinist.com/304/Issue/Article/False/87439/Issue

Chapter 38
Surviving the Data Storm Using Rich Data Structures at Data Recording and Data Warehouse

P. Johnson

Abstract Data collection and analysis for machinery condition monitoring has been completely revolutionized due to the advances of the personal computer (PC). Processors run at GHz speeds, terabyte hard drives at very low cost, and Ethernet networking links systems across the globe. This technology has enabled engineers to perform all types of machinery analysis from thermography to vibration analysis to oil analysis and structural analysis. Engineers are connecting accelerometers, displacement probes, tachometers, cameras, microphones, thermocouples, strain gauges, and a lot more sensors to take the measurements they want. And the result of all this technology is an overwhelming mountain of data that engineers are left to sort through to understand their machinery. This challenge is further complicated by dedicated data collection systems which record just one type of sensory data, such as vibration. Even so, data collection technologies now record data in high fidelity, using 24 bit analog to digital converters, and stream more of it to disk than ever before.

38.1 Introduction—Data Rich and Information Poor

Machine performance data is constantly generated (as long as the machine is operating) whether or not the data contains new information. The decline in cost of data acquisition hardware, the increasing trends in embedded processing speed, and embedded storage capacity have combined to create an explosion of data stored in files and databases. On one hand technology is enabling faster and richer data retention. However, managing and making good use of the data remains a challenge. In today's globally competitive industrial environments, companies need to turn machinery data into usable information in order to maximize uptime and machinery availability.

P. Johnson (✉)
National Instruments, 11500 North Mopac Blvd, Austin, TX 78759, USA
e-mail: preston.johnson@ni.com

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_38,

Several roadblocks can get in the way of optimal exchange of information, the most notorious of which is the improper capture of information initially. All too often the data that is recorded is incomplete or of the wrong type. Data can be incomplete because of choosing the wrong phenomena or not enough types of items to monitor. It may also be imperfect due to limitations in the recording system. If the data is stored without descriptive properties, in inconsistent formats, or scattered about across an arbitrary number of computers, it is just like the data was never collected at all. All of these factors create a graveyard of information that makes it extremely difficult to locate the right data set and derive decisions from it. As a result many companies are seeing a loss in efficiency and not reaping the benefits of online machine monitoring and the information that can result from analysis of properly collected, formatted, and described data.

Presuming one captures good data with sufficient descriptive attributes; the next hurdle is making the data available for maintenance and operations decisions, or in other words asset management functions. Solutions to this challenge range from primitive file and folder naming conventions to complex database solutions that are costly to set up and maintain. The ability to turn raw data into interpretable results instantaneously and to easily share these results throughout the organization is critical for the efficiency of an online machinery monitoring program.

To realize an effective predictive maintenance program, proper choices regarding data acquisition, and data management need to be primary considerations.

38.2 Seamless Exchange of Information

The situation described above scales from a single machine to global facilities. Predictive maintenance goals are generally broken down into a certain number of manageable subtasks by machine or system. For instance, consider the goal of an automaker to improve the fuel consumption of a vehicle. This can be accomplished by individually optimizing and improving the body, engine, transmission, chassis, tires, and so on. However, the best results will come from the "interplay" of all the components of the car. Meeting goals such as increased uptime or more efficient asset utilization requires the same "interplay" from the entire facility, not only a single machine.

Maintenance engineers, like all engineers, want the maximum amount of data when making decisions. To get the most thorough picture of a machine's health, you would need to measure everything and anything you can, but obviously this isn't ideal. Time, cost, and other factors limit us to a subset of data. However, the opposite, to only monitor one variable or otherwise monitor the same list of variables for every machine isn't ideal as well. That it is why it is critical that predictive monitoring equipment be able to measure everything from vibration to thermal imaging to ultrasonic to power quality. An engineer using a flexible modular monitoring device is not constrained by the capabilities of the system and can choose a list of appropriate variables to measure.

38.2.1 Smart Data Recording

One compromise to the “measure everything” mentality is to analyze raw sensory data at the data collections source. If there is not any new information, then do not store or transmit the data, other than to report general machinery status. In other words, one reserves communications bandwidth and disk space for those high fidelity recordings that have new information, or perhaps a daily snapshot of the machine or system behavior. State of the art data recording devices offer modular input configurations that support most analog sensors and even digital sensors. The state of the art data recording device is able to adapt to those specific sensors that have the best indication of asset health. Further, the state of the art data recording device has powerful processing to filter, reduce, store, and forward data that carries information.

Figure 38.1 depicts the embedded data recording analysis process for filtering and reducing data. In this architecture, the embedded data recording device continually collects and analyzes sensor data to determine if there is new information in the data. These information triggers cause a data recording of the appropriate length to be created. Further, the embedded data recording device often updates a remote user interface in the case a remote display is enable to displace current machine status.

38.2.2 Visualizing Information Flow

To better visualize information flow from the data recording device, consider Fig. 38.2. In this modular data recording architecture, analog, and digital input modules are inserted into a modular backplane architecture. These modules perform the function of collecting sensor data for the embedded computing function to analyze, communicate, and store as necessary.

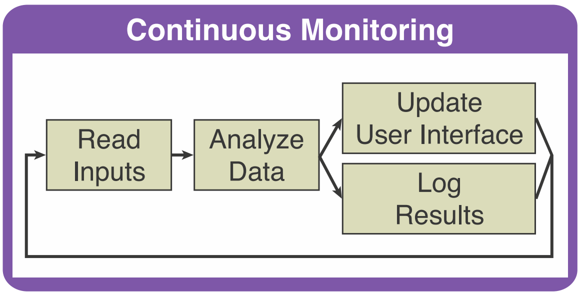


Fig. 38.1 Data recording architecture for filtering and analyzing sensor data

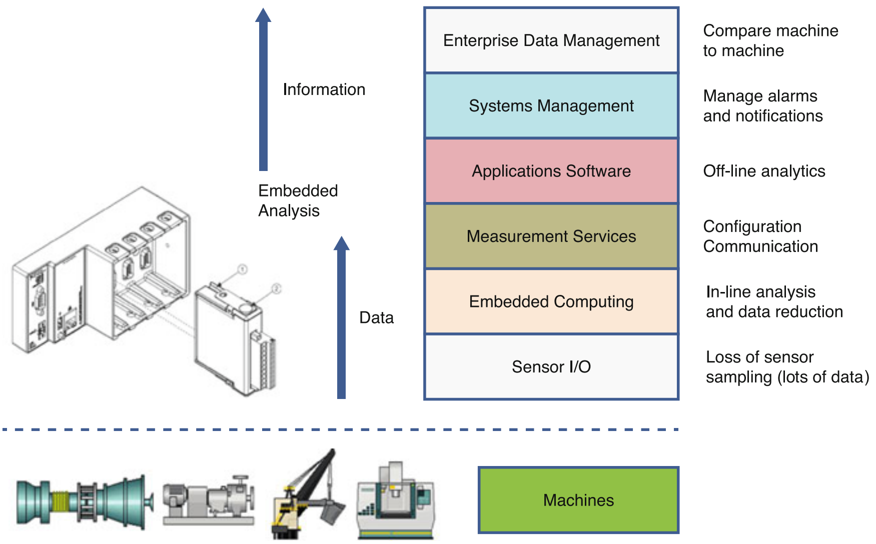


Fig. 38.2 Modular and intelligent data recording architecture

Measurement services provide the communications and the library of analysis tools which the embedded data recording device relies on to analysis and communicate. Once data with new information is identified, it is stored, and forwarded to an application software data service on a data base machine or network. With systems and enterprise data management, data and information from like machines can be compared for improved asset management functionality.

For example, consider a condition monitoring data recording application within a swing shovel used in surface mining applications. Accelerometers are used to record vibration data at roller bearing and gear points within the shovel. The best data is collected at operational cycles when the shovel motors are rotating at relatively high speed while under load. A good data capture is required just once per day. The embedded computing component of the Fig. 38.2 architecture is able to reduce high bandwidth data, store it locally, and determine the best data set to communicate to the maintenance team. The historical alternative of streaming high bandwidth raw accelerometer data over the mining site wireless network is not possible.

38.3 Data Storage Format Considerations

When considering data storage formats, it is best to leverage a technology or format that works well for the embedded data acquisition system, and provides rich descriptive capabilities for downstream prognostics analysis. One common “schema” for data recording is the CHRIS schema as defined by the Machinery

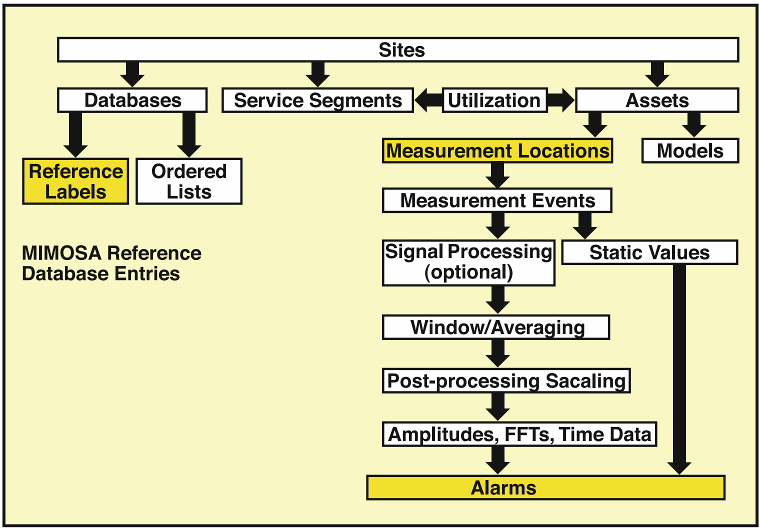


Fig. 38.3 MIMOSA CHRIS schema

Information Management Open Systems Alliance (MIMOSA) organization, MIMOSA [1]. This schema defines data types including time waveform data, spectral data, alarm status, process data, and a range of sensory source information including machine asset and sensor asset information. An illustration of the MIMOSA schema is given in Fig. 38.3.

The MIMOSA data schema describes a rich data architecture allowing for a combination of time waveforms, spectrum, scalar values, images, and related data types to be stored in a unified data base. When the data sets are organized by sensor, mechanical component, etc., a view of related data sets is easily obtained. For example, opening a sectional view under a roller bearing, one would see time series vibration data, temperature trends, vibration spectra, oil particulate count

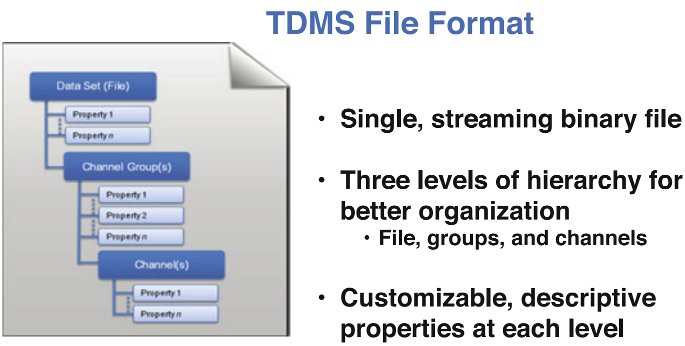


Fig. 38.4 Example embedded data recording structure

trends, etc. All of the information is organized as sensory information related to the bearing.

There are several ways to implement an embedded data storage capability which supports this rich data structure. These include common relational database structures and data structures specifically designed for embedded monitoring applications. An example of a embedded data structure format is shown in Fig. 38.4.

Figure 38.4 illustrates a data structure that is efficient in recording with its high speed streaming capabilities. It is rich in data descriptors with the use of data property strings for each data element or channel stored in the data file. In other words, information about the sensor, location, scaling factors, filtering, and mechanical component beneath the sensor, can be stored as labels that describe the time waveform recording. In addition, properties in the data file describe the conditions that caused the data file to be recorded, whether it be an analysis result threshold limit, a time limit, speed change, or operator request. In addition, this particular implementation requires little support from formal information technology teams, other than to provide the hard drive space to store the measurement data files generated from the embedded data recording devices.

A second feature of this data structure is the ability to add information along the prognostic system or asset management information process. In other words, as the data file record is moved from the data acquisition device downstream to an engineering workstation, additional analysis can be performed on both time series data and extracted features which are stored alongside the original sensory data record, Fig. 38.5.

The task of analyzing and categorizing data is rarely complete. With a flexible data source, additional analysis results, methods, comparisons, labeling, etc., can be added to the data set at any time in the progression of the prognostics process. Going further, if specific empirical patterns emerge, the data files become new models or fault mode references.

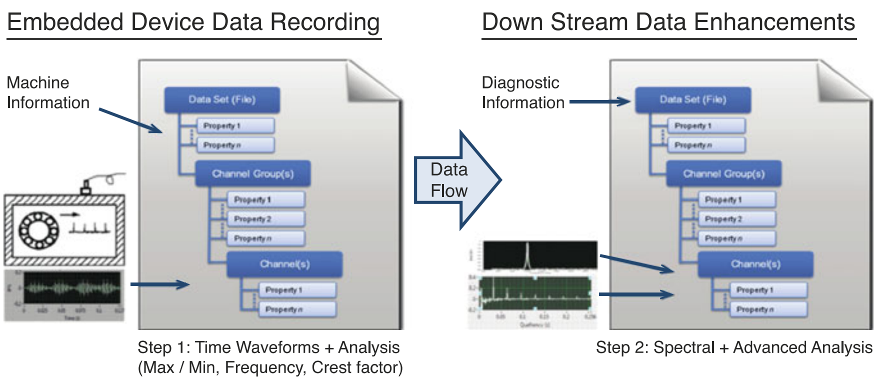


Fig. 38.5 Progression of data structure

38.4 Managing and Mining Data and Information

Even with consistent data formats containing descriptive information, the challenge still remains to make the data available for operations, management, and engineering to both easily access and effectively use. How do you ensure that all data consumers in the organization can find needed data? Common approaches to these challenges range from primitive file and folder naming conventions to complex and costly database solutions. The ability to turn raw data into interpretable results and easily share these results throughout the organization is critical to the efficiency of the asset management process [2].

Many engineers tolerate these challenges because this data is easily loaded into readily available spreadsheet tools such as Microsoft Excel. However, the cost saved by not having to purchase dedicated engineering software is soon lost when time and resources are spent recreating common engineering analysis and reporting routines unavailable in business-centric spreadsheet packages and manually searching for data.

At the opposite end of the spectrum is the approach of using dedicated databases for managing technical data. Because of the reputation for organizing data for easy searching and retrieval, databases have garnered the attention of many test engineers as one way to manage test data. There are some obvious and some not-so-obvious drawbacks to adopting a standard database such as Oracle or Access for your data management system. First, these databases are not set up to work out of the box with high fidelity test data; there is a significant amount of data model and database design work that must take place up front. In addition is the often overlooked ongoing maintenance required to keep databases up and running. This maintenance and scaling over time not only costs more money but usually requires the engagement and dedication of information technology experts, who understand business information well, yet asset management high fidelity data less.

In the end, neither the file and folder approach nor the database approach meets the needs of most test groups. While they may seem like viable solutions at design

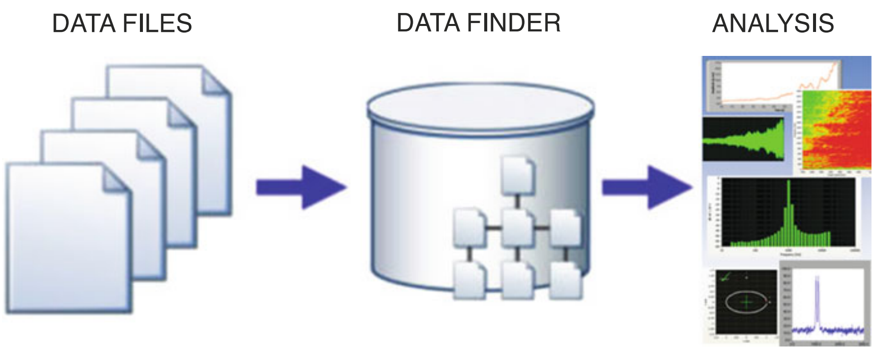


Fig. 38.6 Data management solution using rich and flexible data files with automatic indexing

time, you are still left with maintenance headaches and inefficiencies as your test needs change over time and you continue to store more data.

What is needed is a data management solution to solve these data management challenges and make it easier and faster for information consumers to move from data collection to usable results. The key element is a data mining and data finding tool that leverages the descriptive properties and calculated features to quickly find and report on data and information in the TDMS data structure shown in Fig. 38.4. A technology does exist entitled “NI Datafinder” that creates a data index from the properties and descriptors stored along with the TDMS data structure. This data index uses a hash table which allows a flexible web like search query. Figure 38.6 summarizes the data architecture.

38.5 Conclusion

By incorporate intelligent smart data recording devices with flexible and modular sensor inputs, data can be filtered, reduced, and stored at the point of data acquisition. The resulting file, (when labeled with signal analysis results, machine name, location, type, and so on) can be indexed in a manner that allows results from differing data files to be indexed and searched as a collection. Data files can be analyzed further at the enterprise level and labeled as empirical models of failure modes for specific machine classes.

References

1. Machinery Information Management Operational Systems Alliance (2002) Overview of MIMOSA. Available from: www.mimosa.org
2. Test Data Management for the Enterprise (2011) National instruments NI developer zone. Mar. Available from: http://zone.ni.com/devzone/cda/tut/p/id/5502

Chapter 39
Visualization Management of Industrial Maintenance Data Using Augmented Reality

D. B. Espíndola, C. E. Pereira, R. V. B. Henriques and S. S. Botelho

Abstract This paper proposes the use of the visualization techniques through Augmented Reality (AR) in order to manage maintenance data visualization. Using this method the AR visualization will insert the human in the predictive maintenance loop. It will facilitate the users understanding (in this case the maintenance operator) and visualization of the information from a predictive system and also represent a way to provide a safe interaction for the operator.

39.1 Introduction

Nowadays, integration in the industrial area means "digital engineering" and efficient information management. This assertion is evidenced by the increasing amount of digital systems, which appear daily in the industrial environment [1, 2, 3]. However, the lack of standards in digital formats, the usability of the communication interface between man and computer; the understanding of the information becomes critical issues with the rapid emergence of these technologies.

On the other hand, the complexity of the industrial processes demands the use of visualization tools in their systems. Moreover the use of computational

D. B. Espíndola (✉) · C. E. Pereira · R. V. B. Henriques
Federal University of Rio Grande do Sul—UFRGS, Porto Alegre, RS, Brazil
e-mail: danubiaespindola@furg.br

C. E. Pereira
e-mail: cpereira@eletro.ufrgs.br

R. V. B. Henriques
e-mail: rventura@ece.ufrgs.br

D. B. Espíndola · S. S. Botelho
Federal University of Rio Grande—FURG, Rio Grande, RS, Brazil
e-mail: silviasb@furg.br

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_39,

softwares and sensors monitoring the use conditions of equipment allows an evolution from the traditional systems of corrective and preventive maintenance to predictive systems [4, 5, 6].

In this sense visualization techniques based on Augmented Reality (AR) arise as a solution for 3D visualization capable of helping in these processes. The use of AR techniques allow the superimposing virtual objects generated by computer in the real environment in real time, represents a powerful support tool for industrial maintenance. For this purpose, visualization devices such as TabletPC, Personal Digital Assistant (PDA), Head Mounted Displays (HMD), and so forth can be used to interaction Human–Machine [7, 8].

This paper focuses on "how to" manage the visualization of large amounts of data from systems, such as: 2D graphs, 3D models, maintenance guides text, video, text file, that can be received by operator during maintenance task. The augmented interface proposed here presents data in two ways: autonomous and by command. Both of them are controlled by integration model describes in XML format and structured by graph theory.

So, the integration between maintenance systems and Augmented Reality techniques give a significant contribution both in the state of the art, and in the industrial context. The search for effective improvement in the visualization process has become fundamental for the requirements of cost, time, and quality process [9, 10]. This can be observed through the growing discussion about the use of Virtual and Augmented Reality in industry applications [11, 12, 13, 14].

This paper is divided into 5 sections: Sect. 39.2 presents the related work about application of Virtual and Augmented Reality techniques in the visualization of industrial maintenance processes. Section 39.3 shows the methodology from modeling to visualization for integration of both technologies (maintenance systems and augmented visualization). Finally, Sect. 39.4 shows some case studies and Sect. 39.5 describes the conclusions.

39.2 Related Work

The Augmented Reality (AR) used in the visualization processes have the capacity of mixing real and virtual elements in an environment presented on the output device in real time. The AR and AV (Augmented Virtuality) are inserted in the context of Mixed Reality. According to Milgram et al. [15], Augmented Reality occurs when virtual objects are brought to the real environment. Augmented Virtuality is the opposite, real objects are brought to the virtual environment.

As the focus of this work is the industrial maintenance and more specifically predictive maintenance, aspects of Augmented Reality will be explored as a way of mixing virtual information with real environment (plant/equipment). The use of augmented visualization will allow analysis, interaction with graphs and exploration of cognitive aspects become the understanding of information easier, and facilitating operators in the decision making during maintenance task.

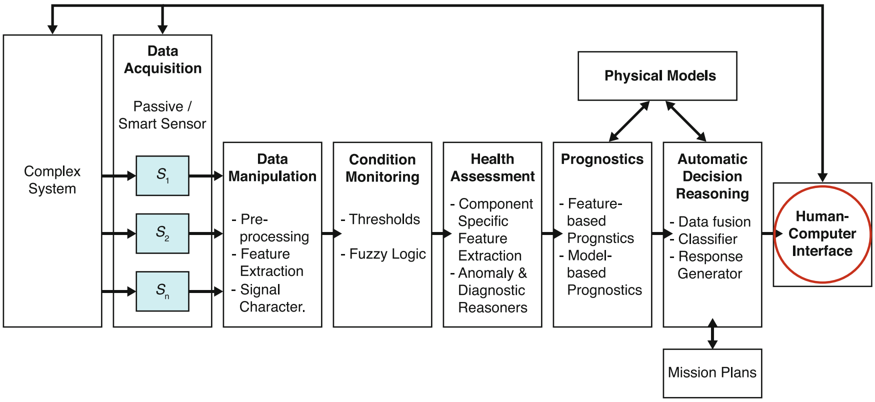


Fig. 39.1 OSA–CBM architecture [16]

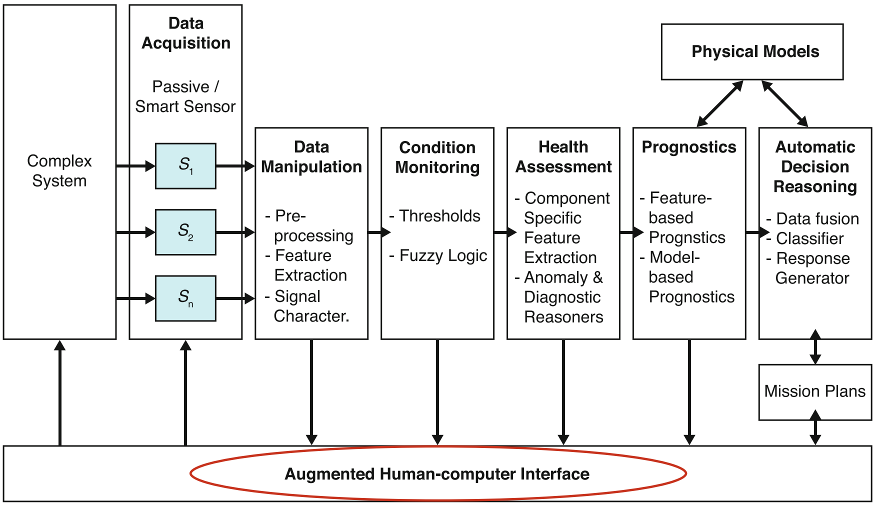


Fig. 39.2 Augmented human–computer interface applied Lebold' architecture

The Fig. 39.1 shows the OSA–CBM architecture proposed by Lebold [16]. During the literature review, was verified that the module *Human–Computer Interface* is few explored in the researches about predictive maintenance. On the other hand maintenance routines and failure diagnosis can be improved through the use of augmented visualization. This work intended to develop a solution for implantation of a visualization system that can indicate and bring information of predictive modules for the user through augmented visualization techniques according to Fig. 39.2.

The augmented reality allows bringing the right information, in the right place on the right moment. Through the use of interaction devices, the operator can request data from all modules of architecture. Besides the AR interface enable the use of video, voice, audio, text, and graphs formats. This makes the interface more intuitive [17].

According to Henderson et al. [18] the augmented reality and predictive systems represent precursory solutions in industrial context [19, 20]. For this reason some events in Virtual/Augmented Reality treat the use of AR/VR in the industry as a "potential application" in the industry [21, 22].

The main applications of AR/VR technologies in the industry are: collaborative project [23], training [20], maintenance [13, 1], and design [21]. Among the initiatives found that using virtual techniques of visualization, there are: ARVIKA [21] and AMIRE [24] projects, STARMATE [22]. More specific research concerning each one of the these areas can be obtained in [25, 26, 27, 28].

39.3 From Project to Visualization

The Fig. 39.3 shows all stages of the information management process to implement augmented Human–Computer interface. A methodology was developed for describes the main stages involved in implementation of proposal. There are four stages in proposed methodology: project, modeling, integration, and visualization. During the project stage were drawn the goals based on requirements and several data models involved in this solution.

This way was developed the high-level architecture to illustrate the data flow, the modules, the services, and the agents that compose the process (Fig. 39.4). Firstly the operator send the command to interface, this command can be by voice or through conventional device (keyboard). The interface is managed by IDM

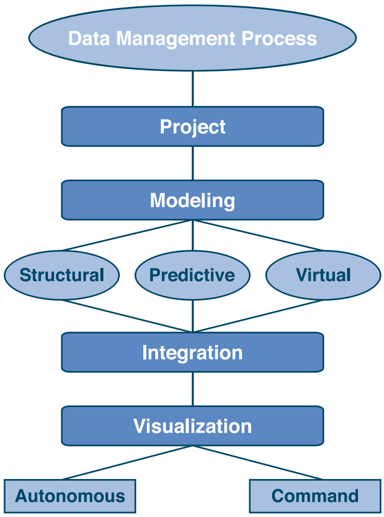


Fig. 39.3 Methodology of proposal

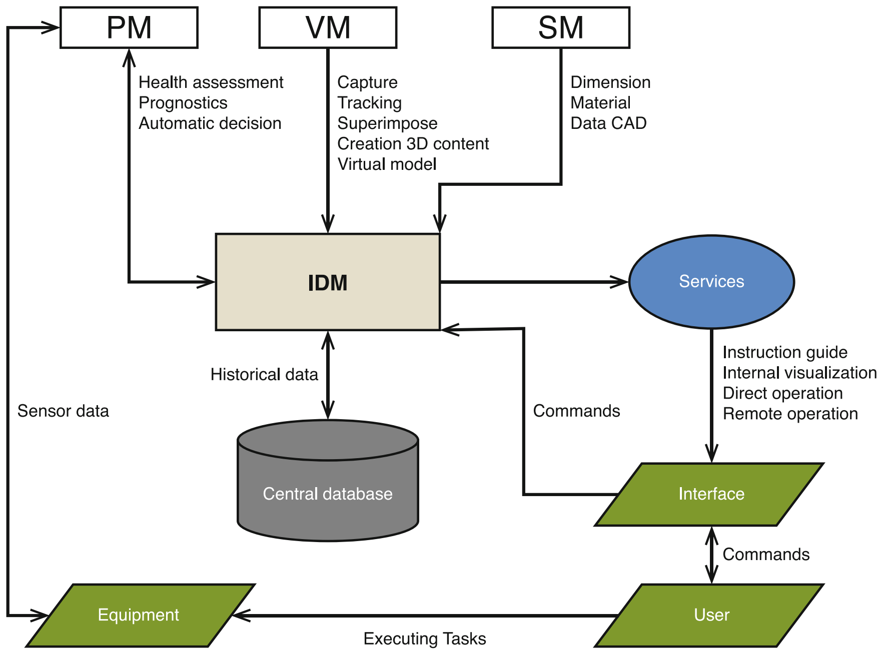


Fig. 39.4 Proposal high-level architecture

(Integrated description model). This model is responsible to search the right information according to user command and present the pertinent virtual service.

The IDM is the core of integration must describe the relationship between the different models involved in the architecture. Several models such as: predictive, structural, and virtual; and information (text, audio, video graph, and voice) to arise from several systems.

According to Fig. 39.4 it can be observed the predictive, virtual, and structural models, which PM, VM, and SM models. These models provide the inputs for the IDM model. The IDM will be responsible for generating a descriptive model in XML format that integrates the different information supplied by the systems and present this information in the augmented visualization interface through services available. Also is possible to view in the high level architecture the data flow of the systems.

The IDM model is the manager of augmented visualization. Therefore after the definition of the architecture in project stage, the next stage was the modeling. Thus, the modeling stage was divided into three parts: structural, virtual, and predictive modeling. Before the modeling the causal analysis was studied to build a failure historical based on data from predictive system, historical data, and expert knowledge supplied by the partner company, in case was CRS Actuator of Coester Company. The cause-effect diagram was implemented to describe the actions to be taken in each case of failure.

The structural modeling refines the model of equipment or plant, i.e., divide the equipment in components hierarchically. This allows describing the equipment in several components and each component in several elements.

The predictive modeling consists into build a model that represents the predictive system outputs for each item of equipment which might fail. This model was described based on cause-effect diagram.

The virtual modeling consists into build a virtual representation of plant/equipment as well as create virtual content such as: virtual text, 3D/2D graph, sound to guide the maintenance tasks. Superimpose 2D pictures as virtual representation of real objects or the conversion CAD model to VR model is a way minimizes the modeling time.

In the integration stage is described in XML format (IDM model) the relationship among PM, VM, and SM models. After this description it is possible to manage the augmented visualization, i.e., shows the information in augmented interface according to the prediction data. The interface operations can be two ways: direct operation (by command) of equipment aided by augmented visualization (virtual instruction guide based on failure case) or remote operation through desktop interface where the operator can have wide vision of maintenance process.

The visualization stage implements two levels: autonomous and command. In the autonomous visualization mode is possible visualize the equipment in general way, this allows check as a whole the operational status of the equipment based on data from the predictive system. In this visualization mode is possible to visualize the instructions about the risk components.

After finalizing the project and modeling, stages the next step was to choose the appropriate tools. Thus, the IMS system was chosen as a solution to the predictive modeling of this proposal. Further details about IMS system can be analyzed in Lee et al. [29]. This system has a core component, called Watchdog Agent [30] that evaluates the working state of machines and equipment.

The Watchdog Agent consists of prognostic algorithms embedded with a group of software tools to predict components and systems failure. This degradation calculation is based on sensor readings which measure the critical properties of the process/machinery. The IMS system supplies a MATLAB tool that allow visualize and process some data. Besides the outputs supplied by Watchdog are: data manipulation, condition monitor, health assessment, prognostics, and diagnostics graph files.

For implantation of virtual modeling (VM) of the interface, the ARToolkit library and the C++ language on Visual Studio platform were used. For the 3D modeling the use of VRML and X3D was chosen, once the current CAD softwares have CAD-RV conversion resources. This will facilitate the task and reduce the time of the 3D virtual modeling.

The IDM model was described in XML format. This language generates a nodes hierarchical model regarding the equipment components. Each node in the XML model contains the virtual, predictive, and structural information supplied by several systems. With that, the interface can organize the information of the several systems of intuitive and explanatory way.

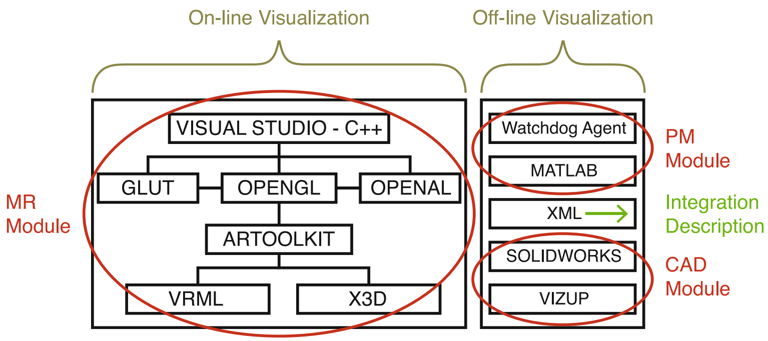


Fig. 39.5 Software architecture

Figure 39.5 represents the software architecture used in the first implementation. The idea is to automate the augmented visualization process through XML description model (IDM) in order to make this process independent of the tools.

39.4 Case Study

The data management process to augmented visualization in maintenance tasks proposed here considers the integration among PM, VM, and SM data. The solutions found in literature that contemplate these themes are rare. However, individual solutions to each one of themes are several, what makes evident the importance these researches.

In this sense some experiments were developed in laboratory to validate the visualization proposal. In this work will be present the CSR Actuator case study. The Fig. 39.6 shows the structural modeling of the case study. The Fig. 39.7 presents the virtual modeling and the Fig. 39.8 shows the predictive modeling.

After the modeling stage, in the integration stage the models presented in Figs. 39.3 and 39.4 are related and described in the IDM model. This descriptive model will manage the data that will be presented at the augmented human–computer interface.

The Fig. 39.9 presents the physical implantation of the PM modeling using the Watchdog Agent. In this experiment was introduced a rubber pipe in the valve entrance to simulate the failure in the opening and closing valve. Moreover information about procedures to be adopted was provided by technicians and inserted into cause-effect diagram.

Along with Watchdog Agent a Matlab toolbox is supplied for the use of prediction algorithms, signal processing, performance assessment, diagnosis and prognostic. With this system, shows in the Fig. 39.8, some graphs were generated for results analysis and integrated in the augmented visualization.

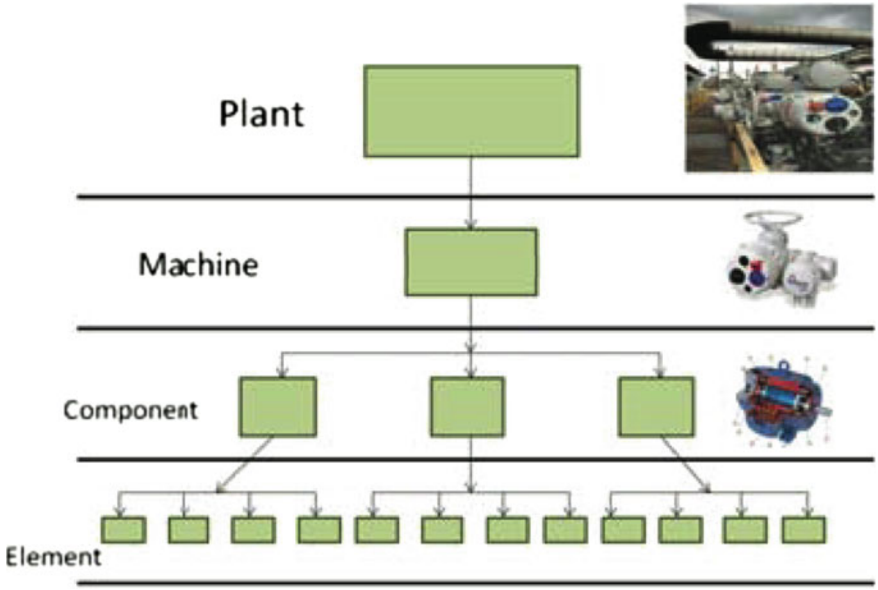


Fig. 39.6 Structural modeling of case study

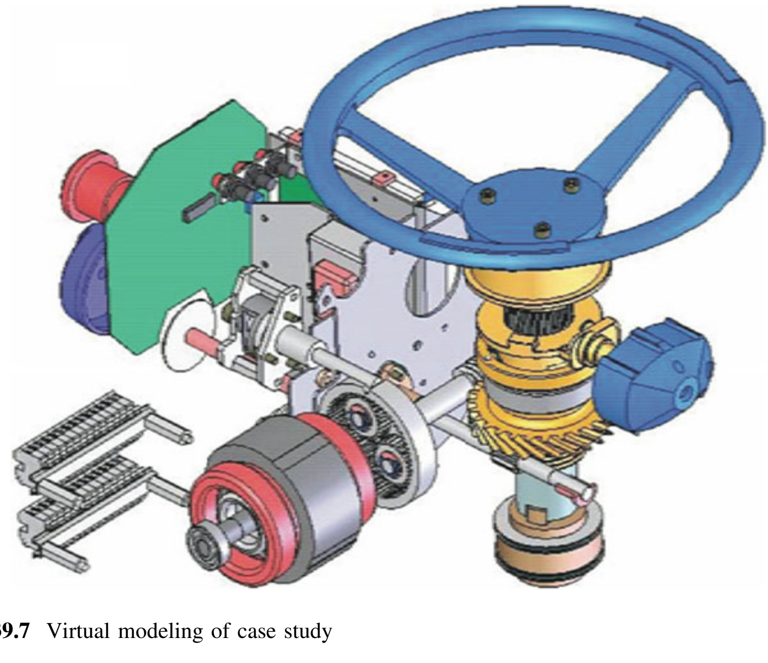


Fig. 39.7 Virtual modeling of case study

Fig. 39.8 Predictive modeling of case study

Fig. 39.9 Layout of physical implantation of the case study

In this case study, the tests were accomplished with desktop device and HMD based on video with coupled camera. In future tests, devices such as: PDAs and TabletPC will be also used. For the processing of the maintenance data and visualization are used a mobile device notebook.

Finally the Fig. 39.10 illustrates the integration among PM, VM, and SM through of the markers. In this figure it is possible observe that each marker corresponds to a element of the virtual model, to an element of the structural model, and an element of the predictive model. It means that each sensor correspond to a kind of marker registered in ARToolkit library, this way the user can access the different components information, activating the corresponding marker.

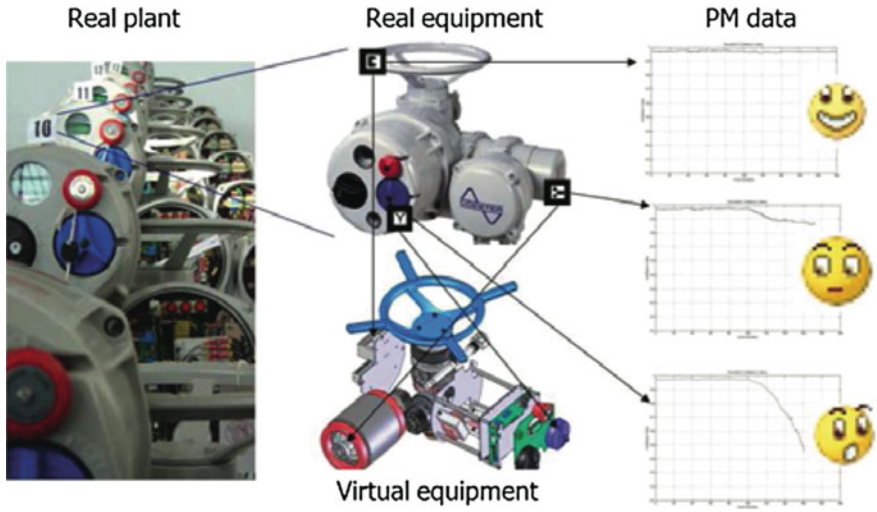


Fig. 39.10 Markers to augmented visualization and data integration

ARToolkit library works as a mixer of virtual and real elements. However, the virtual elements don't appear automatically (at the moment of marker visualization) according the library pattern, but through the request by voice commands, or activation of autonomous visualization by the user.

Several data are captured by sensors, the VM modeling accesses these data in order to search 2D graphs, videos, or texts to represent the sign behavior and indicates the health assessment. The predictive system has graph and text files saved at a standard place from time to time.

39.5 Conclusions

This paper searches the improvement of the maintenance processes through the augmented visualization techniques. The integration among PM, SM, and VM models were fundamental to information management. During the literature review were found several visualization systems for maintenance, however complex and limited. Moreover few studies were found in the "Human–Computer Inteface" area.

A methodology has been proposed for implantation of augmented visualization applied to OSA–CBM architecture, this allows interact with the outputs of each module of predictive systems anytime. After this was proposed a high-level architecture, where was defined which information were involved in augmented visualization solution. The second stage was the modeling processes that defined which models should be integrated.

During the implementation, several software solutions were used as described in the software architecture (Fig. 39.5). In the case study of CRS Actuator the predictive system used (Watchdog Agent) was capable of diagnosing and interpreting the equipment use conditions.

The diagnosis data were transmitted to the augmented interface with the use of ARToolkit library, where the maintenance instructions are presented, as well as 3D virtual information. The information presented in the augmented interface is managed based on cause-effect diagram informed by expert technicians in the maintenance of the valve. These data can help in the process of virtual assembly/disassembly and at the same time, to present maintenance instructions.

It is important highlight that this solution is indicated to critical equipment of industry, where failure can cause great damage by downtime. Therefore, it is necessary to define which are the critical components of the plant, in order to establish the viability in terms of implantation cost of this solution.

In this context, it is not always economically viable the implementation the predictive system. In most of the times it is preferable to replace the component at the moment of the failure [31, 32]. However, the augmented visualization represented by VM modeling proposed is an alternative of low cost where the profits represent a competitive advantage in whatever maintenance system.

References

1. Hassaine S, Dhambri K, Sahraoui H, Poulin P (2009) Generating visualization-based analysis scenarios from maintenance task descriptions. In: Proceedings of the 5th IEEE international workshop on visualizing software for understanding and analysis. pp 41–44
2. Macchi M, Garetti M (2006) Benchmarking maintenance policies in complex production systems. Comput Ind: Spec Issue e-Maintenance 57(6):581–594
3. Muller A, Marquez A, Iung B (2007) On the concept of e-maintenance: Review and current research. Reliab Eng Syst Saf 93(8):1165–1187
4. Lee J, Liao L, Lapira E, Ni J, Li L (2009) Informatics platform for designing and deploying e-manufacturing systems. In: Wang L, Nee A, Yeh C (eds) Collaborative design and planning for digital manufacturing. Springer, Berling, pp 1–35
5. Liao L, Lee J (2009) A novel method for machine performance degradation assessment based on fixed cycle features test. J Sound Vib 356(3–5):894–908
6. Mathew (2006) Reducing maintenance cost through effective prediction analysis and process integration. Adv Vibr Eng 5(2):97–96
7. Stutzman B, Nilsen D, Broderick T, Neubert J (2009) MARTI: Mobile augmented reality tool for industry. Proc World Congr Comput Sci Inf Eng 5:425–429
8. Henderson S, Feiner S (2010) Opportunistic tangible user interfaces for augmented reality. Trans Visual Comput Graph 16(1):4–16
9. Peysson F, et al. (2007) New approach to prognostic systems failures. In: Proceedings of the 17th IFAC world congress
10. You M, Li L, Meng G, Ni J (2010) Cost-effective updated sequential predictive maintenance policy for continuously monitored degrading systems. IEEE Trans Autom Sci Eng 7(2):257–265
11. Regenbrecht H, Baratoff G, Wilke W (2005) Augmented reality projects in the automotive and aerospace industries. IEEE Comput Graph Appl 25(6):48–56

12. Macchiarella M, Vincenzi D (2004) Augmented reality in a learning paradigm for flight and aerospace maintenance training. In: Proceedingas of the 23rd digital avionics systems conference. p 5.D.1–5.1–9
13. Multanen P, Mäkiranta A, Nuutinen P, Leino SP, Helin K, Lind S (2009) Adaptation of virtual tools to support cost-effective maintenance of machines. In: Proceedings of the 22nd international congress condition monitoring and diagnostic engineering management. San Sebastian, Spain
14. Stefánik A, Gregor M, Furmann R, Skorík P (2008) Virtual manufacturing in research and industry. In: proceedings of the 9th Ifac workshop on intelligent manufacturing systems. Szczecin, Poland
15. Milgram P, Kishino F (1994) A taxonomy of mixed reality visual displays. IEICE Trans Inf Syst, E77–D(12)
16. Lebold M, Thurston M (2001) Open standards for condition-based maintenance and prognostic system. In: Proceedings of MARCON 2001—Fifth annual maintenance and reliability conference. Gatlinburg, USA
17. Violante P, Drake K, Latman S, Mandayam S (2008) Virtual environments, data fusion, and video in support of remote health management systems. In: Proceedings of the IEEE international workshop on haptic audio visual environment and their applications. pp 48–52
18. Henderson S, Feiner S (2009) Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In: Proceedings of the IEEE international symposium on mixed and augmented reality. pp 135–144
19. Schoenfelder R, Schmalstieg D (2008) Augmented reality for industrial building acceptance. In: Proceedings of the IEEE virtual reality conference. pp 83–90
20. Schwald B, Laval B (2003) An augmented reality system for training and assistance to maintenance in the industrial context. Proceedings of the 11th international conference in central europe on computer graphics, visualization and computer vision. pp 425–432
21. Friedrich W, Jahn D, Schmidt L (2002) ARVIKA–Augmented reality for development, production and service. In: Proceedings of the international symposium on mixed and augmented reality. p 3–4
22. Schwald B, Figue J, Chauvineau E et al (2001) STARMATE: Using augmented reality technology for computer guided maintenance of complex mechanical elements. E-work and ECommerce 1:196–202
23. Foursa M, d'Angelo D, Narayanan R (2008) Innovative control system for industrial environment. In: Proceedings of the fourth international conference on autonomic and autonomous systems. pp 82–87
24. Grimm P, Haller M, Paelke V, Reinhold S, Reimann C, Zauner J (2002) AMIRE–authoring mixed reality. Proceedings of the first IEEE international augmented reality toolkit workshop. Darmstadt, Germany
25. Song J, Jia Q, Sun H, Gao X (2009) Study on the perception mechanism and method of virtual and real objects in augmented reality assembly environment. In: Proceedings of the 4th IEEE conference on industrial electronics and applications. pp 1452–1456
26. Georgel P, Schroeder P, Navab N (2009) Navigation tools for augmented cad viewing. IEEE Comput Graphics Appl 29(6):65–73
27. Fumagalli L, Elefante D, Macchi M, Iung B (2008) Evaluating the role of maintenance maturity in adoption of new ict in the process industry. In: Proceedings of the 9th IFAC workshop on intelligent manufacturing systems. Szczecin, Poland
28. Blum T, Heining S, Kutter O, Navab N (2009) Advanced training methods using an augmented reality ultrasound simulator. In: ISMAR '09 Proceedings of the 8th IEEE international symposium on mixed and augmented reality. pp 177–78
29. Lee J, Ni J (2004) Infotronics-based intelligent maintenance system and its impacts to closed-loop product life cycle systems. In: proceedings of the international conference on intelligent maintenance systems. Arles, France

30. Djurdjanovic D, Lee J, Ni J (2003) Watchdog Agent, an infotronics-based prognostics approach for product performance degradation assessment and prediction. Adv Eng Inform 17(3–4):109–125
31. Bin Lu, Durocher D, Stemper P (2009) Predictive maintenance techniques. IEEE Ind Appl Mag 15(6):52–60
32. Liao L, Wang H, Lee J (2008) Reconfigurable Watchdog Agent® for machine health prognostics. Int J COMADEM 11(3):2–15

Chapter 40
Modeling of Age-Dependent Failure Tendency from Incomplete Data

P. Hagmark and J. Laitinen

Abstract This paper addresses modeling of age-dependent failure rates from incomplete data that includes interval-censored failure ages. Two estimators for cumulative failure rates are presented: a simple non-parametric estimator and a maximum-likelihood method based on the gamma distribution and the non-homogeneous Poisson process. The maximum-likelihood fit of familiar parametric models (e.g., the power law) to the available field data from an aircraft component was far from satisfactory, so a special three-parameter model function had to be worked out. The maximum-likelihood estimate obtained is then used for repeated random generation of different data sets akin to the field data. This way the effect of data set size, censoring rate, and randomness on the non-parametric estimate can be analyzed to get practical appraisals.

40.1 Introduction

Accurate and systematic data is needed for reliability studies of devices. The data is usually collected by operators and maintenance personnel. Because of unsatisfactory and incoherent reporting requirements and facilities, or simply because of incomplete or lost information, the quality may be poor and there can be gaps in the data sets. Therefore the extremely important reconstruction of the history of individual components can be impossible or require too much preliminary time-consuming work. In the worst case, where the original data amount is small and suffers from missing information, the outcomes and conclusions can be downright misleading.

P. Hagmark · J. Laitinen (✉)
Department of Mechanics and Design, Tampere University of Technology, Tampere, Finland
e-mail: jouko.laitinen@tut.fi

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_40,

In this paper, the failure data will be called *complete*, if (a) it includes every failure age, i.e., the cumulated execution time of the unit when the failure occurred, (b) it includes the age at the start and the end of every follow-up interval of a unit, and (c) all follow-up intervals of the units together cover the entire age period to be studied. Our field data, which concerns an aircraft Air Turbine Starter Valve (ATSV), has been stored in a database during the lifespan of 142 units and is slightly incomplete. Although the collecting system does not compel explicit recording of exact failure ages, it was possible to assess 94 % of them in some way. All other failures (6 %) are interval-censored, i.e., localized between two known age points (Table 40.1).

In the case of an interval-censored failure, the length of the two service sojourns around the unknown age point are partially lost and often also skipped in practical calculations. This is not a big problem if one can assume that all failure-ended service sojourn lengths are from the same distribution, but we are going to study *age-dependent* failure tendency, so interval-censoring must be compensated with better theory. Age-dependent failure tendency is usually expressed by the cumulative failure rate. For the estimation of cumulative failure rates we introduce *two methods*, which utilize all information in incomplete data of the type described above. (1) The nonparametric estimator is based on the natural assumption that the position of a censored failure age is pure random in the censoring interval (Sect. 40.4.1). (2) The maximum-likelihood estimation method again employs comfortable and basic properties of the gamma distribution and the non-homogeneous Poisson process (Sects. 40.2, 40.3). (For basic information on stochastic processes, we refer to [1]).

In maximum-likelihood estimation, the choice of the parametric model is critical. The two-parameter power law with well-established related statistical tests is a widely used model [2, 3], but it is inadequate for our field data as a comparison with the non-parametric estimation shows (Sect. 40.4.2). Instead, we present a three-parameter model with practically interpretable features (Sect. 40.4.3). The maximum-likelihood estimate within this model will thereafter represent the real failure tendency behind the field data. By simulating pseudo-real random data sets of different size and with different censoring probabilities we can study two questions of practical importance (Sect. 40.5). What is the effect of incompleteness (censoring) on the non-parametric estimate and how accurate is the non-parametric estimate when the data set is relatively small? This also supports assessing the accuracy of the parametric model built.

Table 40.1 The realizations of four units selected from the ATSV field data (in total 142 realizations)

Unit ID	Start	1st	2nd	3rd	4th	5th	6th
0009	0	?[a]	363	583	1230c[b]		
0023	0	113	393	394	515	856c[b]	
0078	0	?[a]	377	569	?[a]	1251c[b]	
0136	0	?[a]	?[a]	367	?[a]	691	1490c[b]

[a] Denotes an interval-censored failure

[b] Denotes the end of the follow-up

40.2 A Model for Failure Point Processes

In the theoretical parts of this paper (Sects. 40.2, 40.3), the term 'unit' stands for any equipment intended to execute a certain task, and a 'failure' is an unpleasant event that interrupts the task execution. The 'age' of a unit means the cumulated execution time or some other suitable measure of usage, so age does not include down time (for maintenance or other stops).

Suppose a unit has started a service sojourn at the age $a \geq 0$ and let X denote the age at which the next failure occurs. The cumulative distribution of X is assumed to be of the form

$$G(x, a) = 1 - e^{\Lambda(a)-\Lambda(x)}, a < x < L, \tag{40.1}$$

where $L \leq \infty$, and the *model function* $\Lambda(x)$ is continuous and increasing with $\Lambda(0) = 0$ and $lim_{x\to L}\Lambda(\mathrm{x}) = \infty$. The corresponding quantile function (inverse of G) and the related so-called inverse transform for variate generation of X are now given by

$$G^{-1}(u, a) = \Lambda^{-1}(\Lambda(a) - \ln(1 - u)),\ 0 \leq u < 1, \tag{40.2}$$

$$X = \Lambda^{-1}(\Lambda(a) - \ln(U)), \tag{40.3}$$

where U is uniformly distributed on the unit interval. (Note that $U = 1 - U$ as random variables).

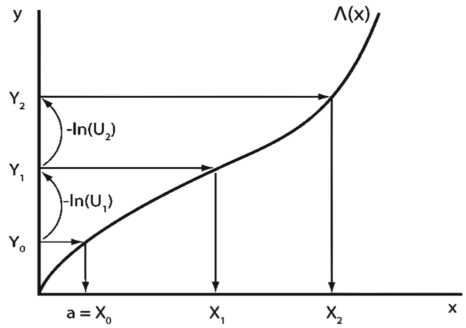
A failure process can be defined as follows. It starts with $X_0 = a \geq 0$, and constant repetition of (40.3) produces consecutive failure age points X_i:

$$X_{i+1} = \Lambda^{-1}(\Lambda(X_i) - \ln(U_i)), \tag{40.4}$$

$$\Lambda(X_{i+1}) = \Lambda(X_i) - \ln(U_i), \tag{40.5}$$

where the U_i are independent (Fig. 40.1). According to (40.5), the Y-process $Y_i = \Lambda(X_i)$ is a renewal process with independent exponentially distributed steps, $-\ln(U_i)$. Thus, the Y-process is a homogenous Poisson process (HPP), and the *failure process* X_i is a non-homogenous Poisson process (NHPP). (For more about generation of Poisson processes, see e.g. [4]).

Fig. 40.1 Illustration of the failure point process



Further, the execution *time to next failure*, $T = X - a$, obeys the distribution $F(t, a) = G(t + a, a)$, so the mean time to next failure and the related variance are

$$MTNF(a) = \int_0^1 \Lambda^{-1}(\Lambda(a) - \ln(u))\,\mathrm{d}u - a,, \tag{40.6}$$

$$VTNF(a) = \int_0^1 \left(\Lambda^{-1}(\Lambda(a) - \ln(u)) - a\right)^2 \mathrm{d}u - MTNF(a)^2 \tag{40.7}$$

Note also that (40.1) immediately implies $\Pr(X > x | X > b) = e^{\Lambda(b) - \Lambda(x)}$ for $b > a$. In other words, a unit that has a stop at age b behaves thereafter statistically in the same way as a working similar unit of age b.

40.3 Interval-Censored Data and its Likelihood

We describe our field data in detail (Table 40.1). It is assumed that the follow-up intervals of all units in the data together cover the age period to be studied. A unit can have several follow-up intervals and the follow-up intervals of different units can overlap. A realization of a unit is a follow-up interval with the failure ages in chronological order; e.g., $X_0 < ? < X_2 < X_3 < ? < ? < X_6 < \cdots < X_n \le X_{n+1}$. The start X_0 and the end X_{n+1} are assumed to be known ages and the index $i = 1, 2\ldots n$ refers to the failures in the open-closed interval $(X_0, X_{n+1}]$. A capital X_i means that the age at the ith failure is known (non-censored), and the ?-symbols (replacing X_i) between two consecutive known age points count to the number of interval-censored failures. Further, the information from a realization can always be split up in *bouquets*, i.e., in sub sequences of the type

$$X_i?\ ?\ldots ?X_{i+k+1}, 0 \le i \le i + k \le n, \tag{40.8}$$

where k interval-censored failure ages are flanked by two known age points, X_i and X_{i+k+1}. Next we formulate the likelihood of the bouquets under the failure process (40.4) with the model function $\Lambda(x)$, Sect. 40.2.

Case $i + k < n$. This refers to a bouquet that is not the last of the realization, so the right-hand limit X_{i+k+1} belongs to a failure. Thereby, the entire step from Y_i to Y_{i+k+1} is the sum of $k + 1$ independent exponentially distributed steps $-\ln(U)$; see Fig. 40.1 and (40.5). Repeated $k + 1$-fold convolution shows that this sum is Gamma distributed with density $e^{-s}s^k \big/ \Gamma(k+1)$:

$$S = Y_{i+k+1} - Y_i = -\sum_{j=1}^{k+1} \ln(U_j) = \mathrm{Gam}(k+1). \tag{40.9}$$

Hence, the likelihood is proportional to

$$e^{\Lambda(X_i)-\Lambda(X_{i+k+1})}\frac{\Lambda(X_{i+k+1})-\Lambda(X_i)^k}{k!}\Lambda'(X_{i+k+1}), \tag{40.10}$$

where $\Lambda'(x)$ is the derivative of $\Lambda(x)$.

Case $i+k=n$. This refers to the last bouquet of the realization. If $k=0$, the bouquet is $X_n = X_{n+1}$ and can be ignored since its likelihood is 1. If $k>0$, we have a sum of two independent related Y-steps: The first step from Y_i to Λ is Gam(k), and the second step, $-\ln(U)$, is censored at Y_{n+1} (the next failure is still to come). Thus, the related likelihood is the convolution of the densities $e^{-s}s^{k-1}\big/\Gamma(k)$ and e^{-s}:

$$\int_0^{Y_{n+1}-Y_i} e^{-s}\frac{s^{k-1}}{\Gamma(k)}e^{s-Y_{n+1}+Y_i}\,\mathrm{d}s = e^{\Lambda(X_i)-\Lambda(X_{n+1})}\frac{(\Lambda(X_{n+1})-\Lambda(X_i))^k}{k!}. \tag{40.11}$$

The total likelihood L is the product of the likelihoods of all bouquets, (40.10) or (40.11), for all units. The variables of L are the parameters of the model function $\Lambda(x$. Finally, the maximization of L provides the optimal parameter values and the maximum-likelihood estimate (MLE), $\Lambda(x)$.

40.4 Modeling of Age-Dependent Failure Tendency

The model function $\Lambda(x)$, Sect. 40.2, obeys an informative and practical interpretation: $\Lambda(x)$ is the *average number of failures in the age period* $(0, x]$. Indeed, it follows quickly from an argument similar to *Case* $i+k=n$ in Sect. 40.3 that the number of failures in a given age interval $(x, x']$ is Poisson distributed with mean $\Lambda(x')-\Lambda(x)$. Hence, $\Lambda(x)$ is simply the cumulative failure rate and the derivative $\Lambda'(x)$ is the failure rate or intensity. We will also use the term 'failure tendency' for both.

In parametric modeling one needs a model function $\Lambda(x)$ that, for a certain choice of parameters, fits the given failure data smoothly. Figure 40.2 depicts an example that is usable for the whole life cycle of many long life systems with several failure modes (e.g. early failures and wear-out failures). However, for our field data this model function would turn out to be unusable. We will soon try two other model functions, but before that we present a natural non-parametric estimate.

40.4.1 Non-Parametric Estimate for Failure Tendency

We record the bouquets (40.8) anew in Table 40.2 for the computational part of our study. The (non-censored) start and end of a bouquet are now denoted by

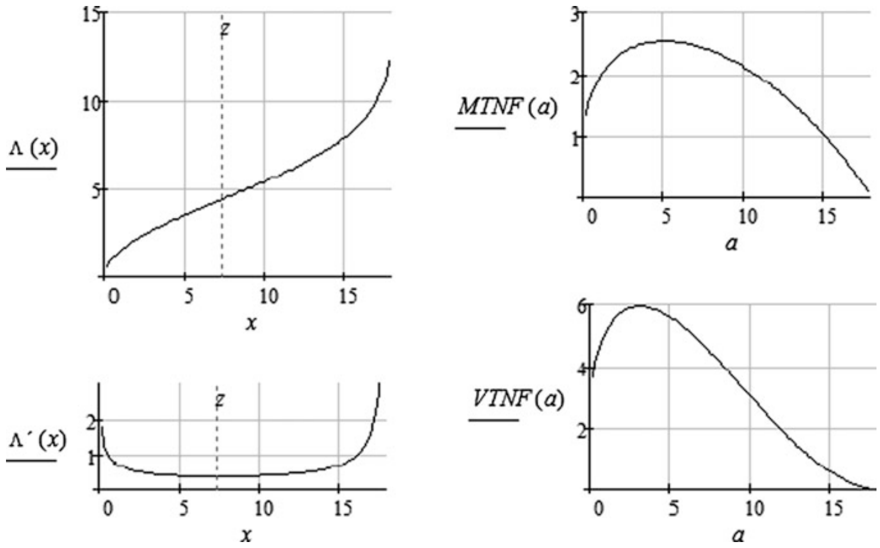


Fig. 40.2 A typical long-term failure tendency for complex systems. [$\Lambda(x) = \alpha(-\ln(1 - x/\gamma))^{\beta}$, $\alpha = 6$, $\beta = 0.476$, $\gamma = 18$, $z = \gamma\left(1 - e^{\beta - 1}\right) = 7.341$]

Table 40.2 Some realizations from the field data in bouquet form (a, k, b, c)

a (start)	0	860	…	0	377	569	…	0	…	0
k (failures)	0	0	…	1	0	1	…	2	…	2
b (end)	860	1,535	…	377	569	1,251	…	367	…	1,777
c (type)	1	0	…	1	1	0	…	1	…	1

$a = X_i$ and $b = X_{i+k+1}$, and k is the known number of interval-censored failures in the open age interval (a, b). The type of the bouquet is $c = 1$ if there was a failure at the end b of the bouquet, and otherwise 0.

Most existing non-parametric methods apply only to complete failure data, where *all* failure ages are non-censored; see e.g. [5]. We present a non-parametric method that is tailor-made for incomplete data of the type described in Sect. 40.3. Let the points $t_0, t_1, t_2 \ldots t_{300}$ be an equidistant sub-division of the entire age period covered by the field data. Let H_v denote the average amount of units and S_v the related amount of failures during the age interval $(t_{v-1}, t_v]$. Using the bouquet format of Table 40.2 we have

$$H_v = \sum_j Lap\left(a_j, b_j, t_{v-1}, t_v\right) \quad v = 1, 2..300, \tag{40.12}$$

$$S_v = \sum_j \left(\frac{Lap(a_j, b_j, t_{v-1}, t_v)}{b_j - a_j} k_j + \left(t_{v-1} < b_j \le t_v\right) c_j \right), \tag{40.13}$$

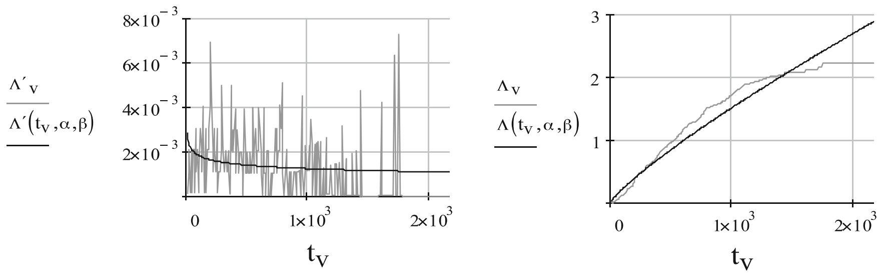


Fig. 40.3 NPE (40.14) and MLE (power law); failure rate and cumulative failure rate

where *Lap*(*a, b, t, s*) is the length of the joint interval (*a, b*)∩(*t, s*), and the expression (*inequality*) equals 0 or 1 according to whether the *inequality* is false or true. The left-hand term in (40.13) expresses the assumption that the interval-censored failure age points in a bouquet are uniformly distributed; the right-hand term again takes care of the possibility that the end of the bouquet is a failure age. Now the non-parametric estimate (NPE) is given by

$$\Lambda'_v = \frac{S_v}{H_v} \;\&\; \Lambda_v = \sum_{r=1}^{v} \frac{S_r}{H_r} \Delta t(\Delta t = t_{300}/300). \tag{40.14}$$

40.4.2 Power Law Estimation of Failure Tendency

As mentioned in the introduction, the power law model $\Lambda(x) = (x/\alpha)^{\beta}$ is theoretically well-established and also widely used in case studies; see e.g. [6] and [7]. The MLE for our field data (Sect. 40.3 & Table 40.2) has the parameters $\alpha = 605.9$, $\beta = 0.831$ and Fig. 40.3 compares the MLE and the NPE (40.14). The difference between the two models for big t-values can very well express randomness, because there are quite few units of age >1500, but the difference in the beginning (more visible in the Λ'-graph) is noteworthy because the early phase is represented by at least 140 units. Clearly, the power law is not tolerable for our data.

40.4.3 A Close-Fitting Model

A model function with only two parameters (like the power law) seldom fits a real failure tendency over long intervals. Mixed and/or piecewise rate models can facilitate the use of standard tests but lead to many parameters. We have set up a

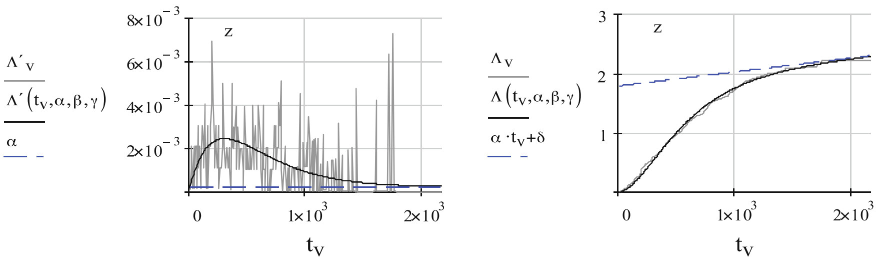
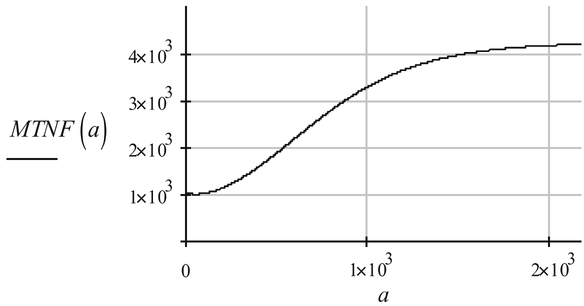


Fig. 40.4 NPE (40.14) and MLE (40.15); failure rate and cumulative failure rate

Fig. 40.5 Mean time to next failure for a-aged units



three-parameter model that is well-defined (non-negative derivative) for all positive parameter values:

$$\Lambda(x,\alpha,\beta,\gamma) = \alpha\left[x\left(1-\gamma e^{-\beta x}\right) + \frac{\gamma-1}{\beta}\left(1-e^{-\beta x}\right)\right]. \tag{40.15}$$

The MLE for our data is achieved for $\alpha \approx 0.000236$, $\beta \approx 0.00338$, $\gamma \approx 26.53$, and the fitting is close, Fig. 40.4. This model also provides some means for cautious practical conclusions and prediction: (1) The failure rate Λ' is maximal at the age $z = (1+1/\gamma)/\beta \approx 307$, and tends asymptotically to α. (2) The asymptote of Λ is $\alpha x + \delta$ where $\delta = \alpha(\gamma-1)/\beta \approx 1.792$. (3) The mean time to next failure tends to a constant: $lim_{a\to\infty} MTNF(a) = 1/\alpha \approx 4.24 \cdot 10^{-3}$. These phenomena can be verified from Figs. 40.4 and 40.5.

40.5 Experimentation with Censoring Rates and Random Data Sets

The true failure tendency behind the field data is of course unknown, but we let it be represented by (40.15), where $\alpha = 0.000236$, $\beta = 0.00338$, and $\gamma = 26.53$ are the maximum-likelihood optimal parameter values (Sects. 40.3 and 40.4.3). Employing this $\Lambda(x)$ in the failure process (40.4) we can simulate complete

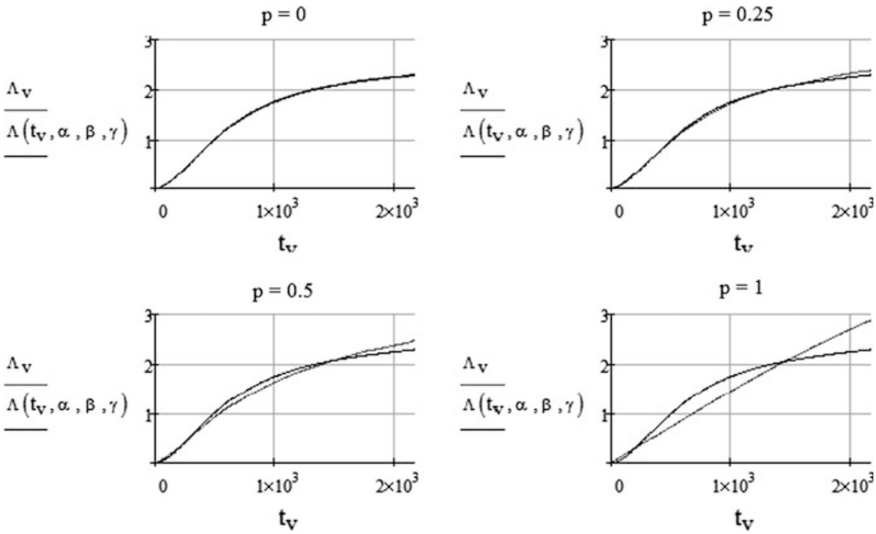


Fig. 40.6 The average impact of the censoring probability on NPE

random data sets. These data sets become incomplete when each generated failure age is censored with a given probability p (cp. Table 40.1). Every data set constructed this way determines of course a unique non-parametric estimate NPE (Sect. 40.4.1).

First, we study the average influence of the censoring probability p on NPE. To minimize the effect of randomness, the data set must be chosen so big that the resulting NPE does not change noticeably when the simulation is repeated. A 40-multiple of the 142 follow-ups in the field data was enough ($40 \cdot 142 = 5680$ realizations). As could be expected, the difference between NPE and $\Lambda(x)$ vanishes for complete data ($p = 0$) and increases with increasing p as Fig. 40.6 shows.

Second, we study how the cumulative NPE reacts on random data sets similar to the field data. From the bouquet format (Table 40.2) we count to $\Sigma c = 263$

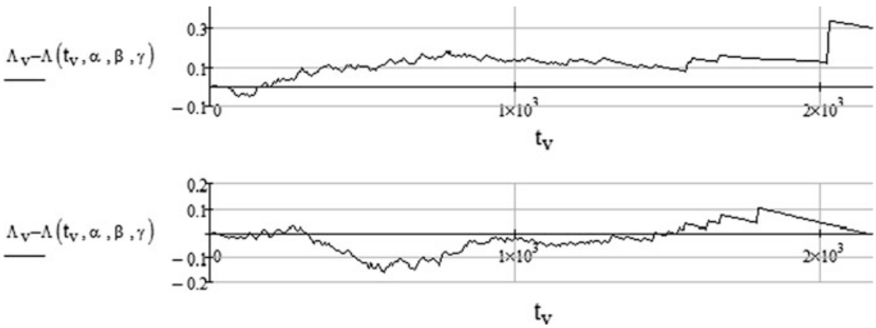


Fig. 40.7 The deviation of two NPE samples

non-censored failures and $\Sigma k = 17$ interval-censored failures, so the fraction of interval-censored failures becomes $\Sigma k/(\Sigma c + \Sigma k) = 17/(263 + 17) \approx 0.061$. We simulated 100 random copies of the field data (the same 142 follow-up intervals) using the censoring probability $p = 0.1$ (could have been 0.061 exactly). Figure 40.7 depicts two examples of how the NPE can vary around $\Lambda(x)$ and the overall average deviation is given by $\Delta = 30000^{-1} \sum_{d=1}^{100} \sum_{v=1}^{300} \left|\Lambda_{v,d} - \Lambda(t_v, \alpha, \beta, \gamma)\right| \approx 0.09$ failures (Sects. 40.4.1, 40.4.3).

40.6 Conclusion

Small-fleet aircraft operators do not have resources to keep a big spare part supply and even the big airlines are facing the same challenge if their fleet consists of small sub-fleets for one aircraft type each. Naturally, if the aircraft fleet is small, the data from devices is meager. Reliability data observed from fielded systems is highly desirable because it accounts for all actual usage and environmental stresses (explicitly or implicitly). Much effort is made to improve failure reporting systems in order to organize and disseminate the data. Despite organizational advancement, field data often lacks sufficient detail. Precise running times to failure are often not recorded in full and only partly comprehensive failure data and operating times are available. Interval-censoring is a typical example: The number of failures in an age interval is known, but the *exact* failure ages are unknown, so the length of two or more service sojourns is partially lost. Undoubtedly, incomplete data must be handled with caution and good theory.

Modeling of age-dependent failure tendency from incomplete data is a two-fold challenge. To overcome the shortcomings of incomplete data of the type described in Sect. 40.3 we have presented two methods for estimation of cumulative failure rates: a non-parametric method and a maximum-likelihood based method. Using a special three-parameter model function we obtain a close maximum-likelihood fit to our aircraft data (Fig. 40.4). Thereafter we show how the average impact of interval-censoring and randomness on the non-parametric estimate can be assessed by simulation (Sect. 40.5). The original censoring rate in our field data is 6.1 %, but we found that rates up to 25 % would have a practically insignificant average impact and that 50 % is still not a catastrophe (Fig. 40.6). This is not surprising since the theory really utilizes the incomplete data effectively. Another practical appraisal follows from the study on randomness (Fig. 40.7): In cases similar to our field data, the average point-wise accuracy of the non-parametric estimate is about 0.1 failures.

References

1. Ross SM (1996) Stochastic Processes, 2nd edn. Wiley, USA
2. Crow LH, Basu AP (1988) Reliability growth estimation with missing data II. In: Annual proceedings of reliability and maintainability symposium. 1988 Jan.; Los Angeles, USA. pp 248–253
3. Antoch J, Jaruskova D (2007) Testing a homogeneity of stochastic processes. Kybernetica 43(4):415–430
4. Pasupathy R (2011) Generating non-homogeneous poisson processes. Wiley encyclopedia of operations research and management science (in press)
5. Arkin BL, Leemis LM (2000) Nonparametric estimation of the cumulative intensity function from a non-homogeneous Poisson process from overlapping realizations. Manage Sci 46(7):989–998
6. Taghipour S, Banjevic D (2011) Trend analysis of the power law process using expectation-maximization algorithm for data censored by inspection intervals. Reliab Eng Syst Saf 96(10):1340–1348
7. Guo H, Watson S, Tavner P, Xiang J (2009) Reliability analysis for wind turbines with incomplete failure data collected from after the date of initial installation. Reliab Eng Syst Saf 94(6):1057–1063

Chapter 41
Asset Lifecycle Information Quality Management: A Six-Sigma Approach

S. H. Lee and A. Haider

Abstract Asset lifecycle management is information intensive. The variety of asset lifecycle processes generate, process, and analyze enormous amounts of information on daily basis. However, as the volume of information increases so does the risks posed to its quality. In engineering asset management, the issue of information quality has organizational, technical, and human dimensions. In technical terms, this issue has its roots in multiplicity of data acquisition techniques, tools, systems, and methodologies, processing of the data thus captured within an assortment of disparate systems, and lack of integration and interoperability. As a result, the information requirements of asset management processes as well as information stakeholders are not properly fulfilled. This paper proposes a novel approach to resolving the issue of information quality. It takes a product perspective of information and applies six-sigma methodology to information quality management. In doing so, it not only assesses the maturity level of the quality of information, but also provides for the continuous improvement of information quality in asset management information systems.

41.1 Introduction

Engineering enterprises mature technologically along the continuum of standalone technologies to integrated systems, and in doing so aim to achieve the maturity of processes enabled by these technologies and the skills associated with their

S. H. Lee (✉)
School of Computer and Information Science, University of South Australia, Adelaide, Australia
e-mail: leesy116@mymail.unisa.edu.au

A. Haider
CRC for Infrastructure and Engineering Asset Management (CIEAM), School of Computer and Information Science, University of South Australia, Adelaide, Australia
e-mail: Abrar.Haider@unisa.edu.au

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_41,

operation [1]. Asset management engineering enterprises have twofold interest in information and related technologies; first that they should provide a broad base of consistent logically organized information concerning asset management processes; and, second that they make available real time updated asset related information available to asset lifecycle stakeholders for strategic asset management decision support [2]. This means that the ultimate goal of using information systems for asset management is to create information enabled integrated view of asset management. This allows for asset managers to have complete information about the assets available to them, i.e., starting from their planning through to retirement, including their operational and value profile, maintenance demands and treatment history, health assessments, degradation pattern, and financial requirements to keep them operating at near original specifications. In theory information systems in asset management, therefore, have three major roles; firstly, information systems are utilized in collection, storage, and analysis of information spanning asset lifecycle processes; secondly, information systems provide decision support capabilities through the analytic conclusions arrived at from analysis of data; and thirdly, information systems provide for asset management functional integration.

Asset lifecycle is information intensive; however, information requirements of lifecycle management processes are prone to change due to the continuously changing operational and competitive environment of asset management. The ability of an organization to understand these changes contributes to its responsiveness to internal and external challenges, as well as its capacity to improve and enhance reliability of asset operations. Thus the real value of information for asset management comes from how effectively information systems are mapped to the asset lifecycle management processes, and how effectively they are synchronized with other information systems in the organization. Even integrated data is of no use, if its quality is compromised. Data quality and its integration in business processes have been handled from a variety of angles in literature (see for example, [3, 4]). Researchers in the past two decades have developed a number of qualitative as well as quantitative information quality (IQ) assessment and management frameworks and methodologies. However, an all inclusive methodology, i.e., the one that assesses as well as enhances asset management information quality, has seldom come to fruition. This paper takes a product perspective of information and presents a six-sigma based information quality management methodology.

41.2 Six-Sigma Methodology for Information Quality Management

Treating information as a product is to provide a well defined product process and produce high quality information product rather than treating information solely as the by-product of business process execution. This is the key idea that this paper

Table 41.1 Products versus information manufacturing [8]

	Product manufacturing	Information manufacturing
Input	Raw material	Raw data
Process	Assembly line	Information system
Output	Physical products	Information products

uses to apply six-sigma methodology to IQ area for quality improvement and IQ assessment. Six-sigma is an organized and systematic method for manufacturing process improvement that relies on statistical methods and the scientific method to make reductions in defect rates [5]. Applying six-sigma methodology to IQ assessment, therefore, provides benefits, such as defining critical factors to quality, measuring current quality (sigma) level, analyzing deficiencies in information and identifying the root causes of poor information, improving quality of information products, and controlling standardized IQ assessment framework. Table 41.1 illustrates the analogy between product manufacturing and information manufacturing to manage information as a product.

Using a product perspective of information, Fig. 41.1 presents a six-sigma we develop an IQ assessment framework. Although this framework appears as assessing IQ, its fundamental core is based on continuous IQ improvement.

41.2.1 Establishing IQ Requirements (Phase 1)

At this stage, the proposed framework seeks information stakeholders' requirements in terms of IQ. These requirements are then translated or mapped to the

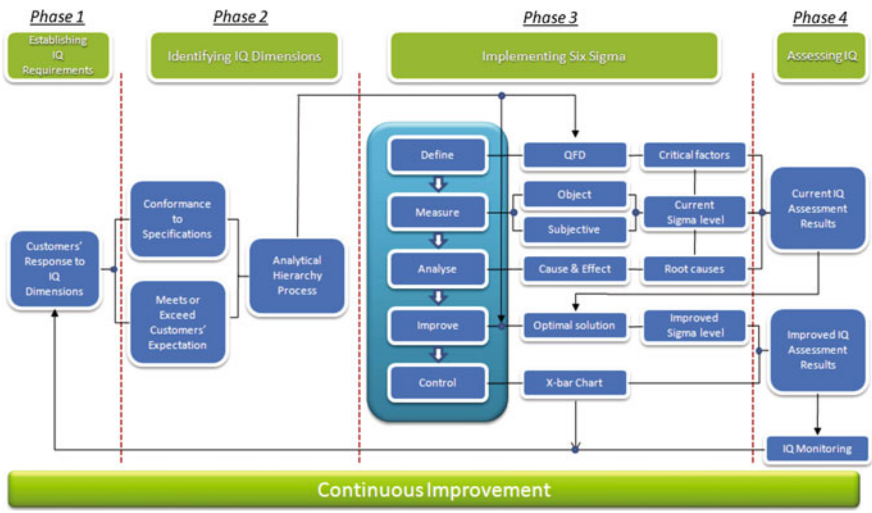


Fig. 41.1 IQ assessment framework by six-sigma approach

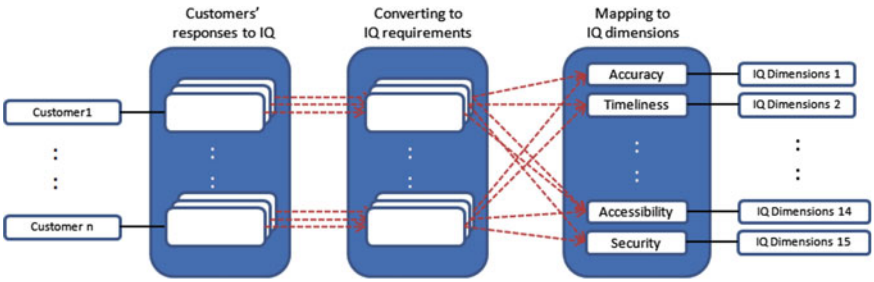


Fig. 41.2 Converting and mapping customer's response to IQ dimensions

various IQ dimensions, such as accuracy, timeliness, completeness, accessibility, and security. The authors propose to survey information stakeholders to collect their responses to IQ (Fig. 41.2).

41.2.2 Identifying IQ Dimensions (Phase 2)

At this stage, IQ dimensions from phase 1 are corresponded to the IQ hierarchy. The IQ dimensions are categorized by the 'conformance to specifications' i.e. product quality; and the 'meets or exceeds customers' expectation', i.e., service quality, respectively. Here, the product quality or information quality implies quality dimensions associated with information, and the service quality include dimensions that are related to the process of delivering right information at right time to right stakeholders. The end result of this exercise is a set of hierarchy or IQ dimensions as shown in Table 41.2. Analytical hierarchy process (AHP) is utilized to correlate these dimensions and assign weights of relative importance. AHP is a hierarchical representation of a system assigning weights to a group of elements by a pair-wise comparison [6]. The pair-wise comparisons operate by comparing two elements at one time regarding their relative importance throughout the whole hierarchy. Therefore, it helps to capture the importance of desired measurement objects in comparison to other objects in the same hierarchy. The assigned weights to IQ dimensions are applied to quality function deployment (QFD) by providing the weights of importance as a scale of importance rating.

41.2.3 Implementing Six-sigma (Phase 3)

This phase applies six-sigma methodology to IQ dimensions, based on DMAIC (Define, measure, analyze, improve, and control). Table 41.3 shows each perspective, procedure, and expected output.

Table 41.2 Hierarchy of IQ dimensions (adopted from [9])

	Categories	Quality perspective	Assessments items
Information quality dimensions for assessments	Conforms to specifications	Product quality	Free of error
			Concise
			Completeness
			Consistence
		Service quality	Timeliness
			Security
	Meets or exceeds customer's expectations	Product quality	Appropriate amount
			Relevancy
			Ease of understand
			Interpretability
			Objectivity
		Service quality	Believability
			Accessibility
			Easy of operation
			Reputation

In order to assess IQ and take advantage of the Six-sigma, this phase follows the method of define, measure, analyze, improve, and control (DMAIC).

41.2.3.1 Define Step

The main objective of this step is to provide a top down view of IQ from a business perspective. Therefore, it relies on,

- Establishing IQ requirements from an information system perspective.
- Profiling information and setting objectives values of IQ.
- Definition of how to measure the quality of information, i.e., defining measurement methods and tools for IQ.

41.2.3.2 Measure Step

The main objective of this step is to acquire data that should lead to calculation of existing level of IQ. It includes,

- Identifying data sources, planning data collection and data sampling.
- Identifying how to interpret information in measurable forms.
- Implementing measurement and calculating the current quality (sigma) level of information in information systems.
- Determining sigma level of available information products.

Table 41.3 Six-sigma perspectives to IQ

Six-sigma phases	Six-sigma perspective	IQ perspective	Procedure	Expected output
Define	Defining the process and customers' satisfactions	Representing customers' requirements (Specification) to IQ and Identifying IQ problems related IQ assessment	Mapping and representing customer's satisfactions to IQ dimensions and requirements	Failure mode and Effects Analysis QFD (Quality Function Deployment)
Measure	Collecting and comparing data to determine problems	Identifying criteria, systems, scales, and scope for IQ measurement. Identifying data sampling method	Implementing measurement of each IQ dimensions and calculating sigma level	Determining current sigma level (Performance)
Analyze	Analyze the causes of defects	Identifying poor information and root cause of poor Information	Critical factor analysis based on measurement results. Applying correlations matrix	Root causes of poor information Fish-born diagram/Logic tree
Improve	Eliminating variations and creating alternative process	IQ improvement by eliminating root cause of poor information	Identify and define specific process improvements for information system	Assessment results IQ assessment framework
Control	Monitoring and controlling the improved process	Standardize IQ	Interpret and report information quality. Document improvement	X-bar Chart for IQ monitoring

- Determining assessment methods (objective or subjective).

IQ assessment can be categorized to objective and subjective measurement. Objective measurement measures information automatically based on its rules or patterns. However, it is difficult to capture the expectations of information stakeholders by comparing them to subjective measurement. Therefore, the defining measurement method should be established in relation to the importance between the ‘conformance to specifications’ and the ‘meets or exceeds customers’ expectations’. The comparison is shown in Table 41.4 represents the comparison of objective and subjective measurement methods and characteristics.

41.2.3.3 Analyze Step

The main objective of this step is to verify and identify root causes of IQ problems. It includes,

- Identifying deficiencies in the quality of information.
- Identifying the root causes of poor information.
- Finding out critical factors of the quality.
- Finding out how information may become deficient in information systems.

A cause and effect diagram is utilized by generating a comprehensive list of possible causes to discover the reason for a particular effect. In this step, a cause and effect diagram is designed based on the results of six-sigma define and measure steps to identify root causes of poor information and critical factors of the quality.

41.2.3.4 Improve Step

The main objective of this step is to identify an improvement of information systems by increasing quality of information products. It includes,

- Seeking the optimal solutions.
- Finding the optimal trade-off values of IQ dimensions.
- Determining whether the information is fit for use in its task.

Table 41.4 Objective and subjective measurement (adopted from [10])

Feature/method	Objective measurement	Subjective measurement
Tool	Software	Survey
Measuring object	Data	Information products
Criteria	Rules, patterns	Fitness for use
Process	Automated	User involved
Assessing result	Single	Multiple
Data storage	Database	Business contexts

- Defining how to establish IQ assessment framework.

All the results from define, measure, and analyze and previous IQ assessment results are integrated to find optimal solutions for IQ improvement. These solutions employ the results of QFD and correlation matrix as well as current IQ assessment results.

41.2.3.5 Control Step

The main objective of this step is to maintain high quality of information. Therefore, it includes,

- Representing IQ assessment for each information system.
- Standardizing the IQ assessment framework.
- Generating documents for IQ monitoring.

Standardization of IQ assessment and monitoring IQ for information systems are required to control IQ. An X-bar chart is a control chart used for monitoring information by collecting sample at regular intervals [7]. Each IQ dimension of sampled information in information system is monitored by using the X-bar chart at regular intervals to prevent production of poor information and to ensure the high quality of information.

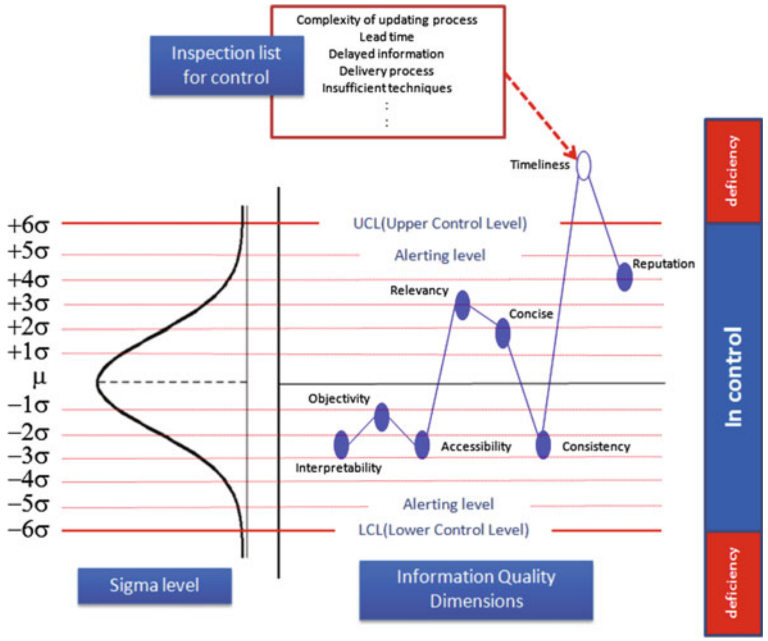


Fig. 41.3 X-bar chart for IQ monitoring

41.2.4 Assessing IQ (Phase 4)

Based on the phase 3, current IQ and improved IQ assessment results are compared to evaluate IQ assessment framework. In order to ensure continuous improvement of IQ, the X-bar chart derived in the phase 3 control step is applied. By using the X-bar chart, if a certain IQ dimension exceeds the specified limits, then the X-bar chart would raise alarm about that dimension.

Figure 41.3 shows an example of IQ monitoring. In this case, "timeliness" exceeds the accepted limits, which indicates that sampled information in an information system is deficient in the "timeliness" dimension. In order to assess IQ, information system is diagnosed with the inspection list which is developed from Six-sigma steps (Phase 3).

41.3 Conclusion

In this paper, we have presented a methodology for IQ assessment and management based on the six-sigma approach. The approach has four major phases, i.e., establishing IQ requirements, identifying IQ dimensions, implementing six-sigma, and continuous assessment of IQ. By treating information as a product, transforming six-sigma perspectives to IQ perspectives, and mapping information on to the six-sigma design, the proposed framework is capable of providing benefits such as, defining critical factors for quality of information, measuring current quality (sigma) level, analyzing deficiencies in information and identifying the root causes of poor information, improving quality of information products, and controlling and standardization of IQ through continuous assessment of data for anomalies. Therefore, this assessment framework leads to accurate, systematic, and pragmatic assessment results.

Acknowledgments This research is a part of the Asset Management Information Auditing and Governance project of the CRC for Infrastructure and Engineering Asset Management.

References

1. Haider A (2007) Information systems based engineering asset management evaluation: operational interpretations (Ph.D. Thesis). University of South Australia, Adelaide, Australia
2. Haider A, Koronios A (2005) ICT based asset management framework. In: proceedings of 8th international conference on enterprise information systems ICEIS, Paphos, Cyprus, 3:312–322
3. Jarke M, Lenzerini M, Vassiliou Y, Vassiliadis P (2003) Fundamentals of data warehouses. Springer, Berlin
4. Naumann F, Rolker C (2000) Assessment methods for information quality criteria. In: proceedings of the 2000 conference on information quality, Cambridge, pp 148–162

5. Linderman K et al (2003) Six-sigma: a goal-theoretic perspective. J Oper Manage 21(2):193–203
6. Cheng EWL, Li H (2001) Analytic hierarchy process: an approach to determine measures for business performance. Meas Bus Excellence 5(3):30–37
7. Hsieh C, Lin B, Manduca B (2007) Information technology and Six-sigma implementation. J Comp Inf Syst 47(4):1–10
8. Wang RY (1998) A product perspective on total data quality management. Commun ACM 41(2):58–65
9. Kahn BK, Strong DM (1998) Product and service performance model for information quality: an update. In: Proceedings of the 1998 conference on information quality, Cambridge, pp 102–115
10. Ge M, Helfert M (2007) A review of information quality research—develop a research agenda. In: proceedings of the 12th international conference on information quality, MIT Press, Boston

Chapter 42
An Ontology-Based Implementation on a Robotic Assembly Line for Supporting Lifecycle Data Management

A. Matsokis and D. Kiritsis

Abstract Issues such as data integration and system interoperability are becoming more crucial in asset lifecycle management (ALM) information models. Improving the availability and exploitation of maintenance data is beneficial for both: predicting malfunctions and failures of the assets; and providing useful feedback of the beginning of life (BOL) to engineers for improving the next generations of the assets. The aim of this work is to combine an ontology information model for semantic maintenance with an IT system architecture in order to provide benefits and new services for the maintenance of the SISTRE system (Supervised industrial system for pallets transfer). The use of the ontology model is combined with description logics (DLs) which allow to reason on classes and instances. The information system is executable, dynamic, and flexible. The use of DLs is tested and validated through implementation in SISTRE which also demonstrates how the user may extend the developed model for facilitating better his needs and at the same time maintain data integration and interoperability among the variations of the initial model.

42.1 Introduction

Issues such as data integration and system interoperability are becoming more crucial in asset lifecycle management (ALM) information models. Improving the availability and exploitation of maintenance data is beneficial for both: predicting

A. Matsokis (✉) · D. Kiritsis
Swiss Federal Institute of Technology in Lausanne (EPFL), STI-IGM-LICP, ME A1 380
Station 9, Lausanne 1015, Switzerland
e-mail: aristeidis.matsokis@epfl.ch

D. Kiritsis
e-mail: dimitris.kiritsis@epfl.ch

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_42,

malfunctions and failures of the assets; and providing useful feedback of the beginning of life (BOL) to engineers for improving the next generations of the assets. The aim of this work is to combine an ontology information model for semantic maintenance with an IT system architecture in order to provide benefits and new services for the maintenance of the SISTRE system (Supervised industrial system for pallets transfer). The use of the ontology model is combined with description logics (DLs), which allow to reason on classes and instances. The information system is executable, dynamic, and flexible. The use of DLs is tested and validated through implementation in SISTRE, which also demonstrates how the user may extend the developed model for facilitating better his or her needs and at the same time maintain data integration and interoperability among the variations of the initial model.

Along the lifecycle of products and assets a huge amount of information is being generated which could be beneficial for the product itself during its usage period; for the new generations of the product; as well as for evaluating the status of the components of the product when it arrives at its end of life. However, this information in many cases is not available due to various integration issues and therefore, it is not analyzed and exploited. Standards such as Machinery information management open systems alliance (MIMOSA) also aim at providing interoperability of different systems and integration of data.

The model used in this paper is named “SMAC-Model” (Matsokis et al. [1]) which is an ontology information model which aims in providing advanced maintenance services as well as feedback for beginning of life (BOL) and input for end of life (EOL). The model is generic and it is developed to facilitate complex physical assets and to work in industrial environment. This model was developed with the aim of being a maintenance-oriented platform providing support for integrating maintenance activities of the middle of life (MOL) and allowing maintenance agents to have remote access to data, to remotely submit requests for diagnosis and to communicate with intelligent systems for better management of maintenance activities. The SMAC-Model is developed using the web ontology language with description logics (OWL-DL) and implements ontology advantages and features which include: data structure layout, description logic (DL) reasoning (Baader et al. [2]), semantic web rules including semantic web rule language (SWRL) [3] and Jess rule engine [4]. With the use of these technologies the model is executable, dynamic, and flexible.

In this work we studied a real lab-scale manufacturing system, the SISTRE system (Supervised industrial system for pallets transfer), and we implemented it in the SMAC-Model. The aim of this work is to combine an ontology information model for semantic maintenance with relative IT system architecture in order to provide benefits and new services for SISTRE in data integration and system interoperability. The structure of the paper is as follows. Section 42.2 describes briefly previous works in e-maintenance, including the SMAC-Model and the system architecture. In Sect. 42.3 there is a brief description of SISTRE system and finally in Sect. 42.4 there is the ontology implementation as well as a demonstration of the final system.

42.2 Background Literature

In this section we briefly describe MIMOSA and its functionality, the SMAC-Model and the IT system architecture. A very interesting review related to e-maintenance works has been carried out by Muller et al. [5]. The authors study standards for e-maintenance, the formalization of e-maintenance processes, the e-maintenance system development and the e-maintenance platform development. The different standards, platforms etc. are positioned according to their contribution to e-maintenance. According to this positioning as a standard it is mentioned MIMOSA and has strong contribution to collaborative maintenance, maintenance data integration, and system interoperability. PROTEUS is strong in remote maintenance, collaborative maintenance, and multi-agent collaboration. PROMISE is strong in knowledge capitalization and knowledge management. The latter two are the basis of the SMAC-Model. Also from this review work it is clear that the definition of what e-maintenance is vague and really depends on each author's point of view. However, the common point of the various definitions is the aim to provide new maintenance services using some form of ICT technology.

42.2.1 Mimosa

MIMOSA [6] is an alliance focused on developing consensus-driven open data standards to enable interoperability of "operations and maintenance" (O&M) processes, systems, and actors. Standards developed in the framework of MIMOSA are aiming to be widely accepted and to be used in facilitating seamless asset management data exchange through integration. For the enterprises adopting such standards the result will be to eliminate the information gaps which exist among the different systems such as real-time control systems and business information systems. The aim of MIMOSA is to encourage the adoption of open information standards by introducing the MIMOSA OSA-EAI (Open system architecture for enterprise application integration). This is a standard for data exchange of engineering asset management data about all aspects of equipment, including the physical configuration of platforms, the reliability, condition, and maintenance of platforms, systems, and subsystems. To this end MIMOSA provides a series of interrelated information standards: the Common conceptual object model (CCOM) which provides the basic conceptual model basis for OSA-EAI; the Common relational information schema (CRIS) which provides a common implementation schema and allows information from many systems to be communicated and integrated (vertical and horizontal integration); metadata reference libraries and a series of information exchange standards which use XML and SQL. Advantages of using OSA-EAI as a basis to build databases for asset management data include software re-use and data interoperability (Mathew et al. [7]). Another MIMOSA standard is the OSA-CBM (Open system architecture for condition based maintenance) which provides the architecture for moving information in a condition-based maintenance system and also provides the tools for implementing the architecture. Its architecture is based on ISO-13374 and comprises of six blocks: data acquisition, data manipulation, state detection, health assessment, prognostic assessment, and advisory generation (OSA-CBM primer). The first three involve only devices which collect and process data to detect abnormalities. The rest combine devices and human agents to define the health status of the equipment, to predict future faults and to support decision. OSA-CBM uses many of the data elements that are defined by the OSA-EAI and in the future the aim is OSA-CBM to be mapped into OSA-EAI.

Moreover, there are some works which propose changes and extensions of the model to make it more generic and to cover a wider area of the domain. Voisin et al. [8] in the framework of FP6 project DYNAMITE: Dynamic Decisions in Maintenance developed an e-maintenance platform to include prognosis in their model. The authors suggested extensions in several parts of the OSA-CBM model in order to formalize the prognosis objects and data. Mathew et al. [7] in their work developed a condition monitoring system called BUDS which claims to support advanced diagnosis and prognosis models not available in commercial systems. BUDS database is based on OSA-EAI and the authors describe several issues of the OSA-EAI which they defined during the development of BUDS including the lack of documentation and excessive normalization.

42.2.2 The SMAC-Model

The SMAC-Model is an ontology information model which aims mainly in providing semantic maintenance. SMAC-Model provides multi functionalities: supports integrating maintenance activities of the MOL, maintenance agents have remote access to data; they can remotely submit requests for diagnosis and they can communicate with intelligent systems for better management of maintenance activities. Moreover, the model supports Closed-Loop Product Lifecycle Management. In this way the data and the information produced from the asset during its MOL is collected and processed to be used as input for improvement of BOL activities (design, production), and EOL activities (recycling, remanufacturing, reuse, etc.). Thus, this model allows closing the information loop between the different phases of the lifecycle. The class diagram of the SMAC-Model is shown in Fig. 42.1. Applications of the SMAC-Model (Matsokis et al. [9]) show the usage of the DLs and the DL-reasoner for re-classifying the class-hierarchy, for discovering the equivalencies among concepts, for checking the consistency of the model, understanding the expressivity of the model and finally, for performing the logical categorization of the instances under the classes. These functionalities support data integration among the different variations of the system and therefore, data is automatically located in the right logical place in the model.

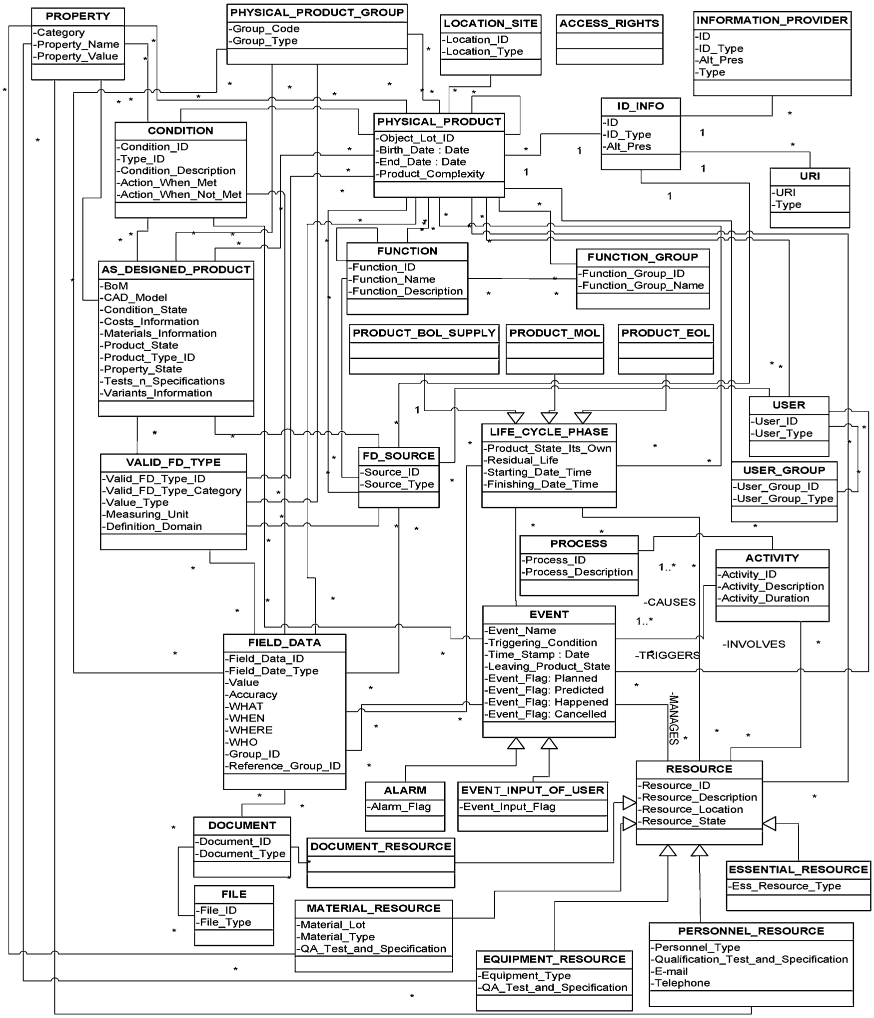


Fig. 42.1 The structure of the SMAC-Model

42.3 The SISTRE System

SISTRE is a pallet transfer system, which represents a flexible production system. It is composed of five robotized working stations, which are served by a transfer system of pallets organized into double rings: inner and outer (Fig. 42.2). The pallets are conveyed on the inner ring, which allows the transit between the various stations. RFID tags (one on each pallet) determine the pallet path through the system. When the pallet has to be handled by a robotic machine in a certain working station, the pallet is deviated on the outer ring where the concerned

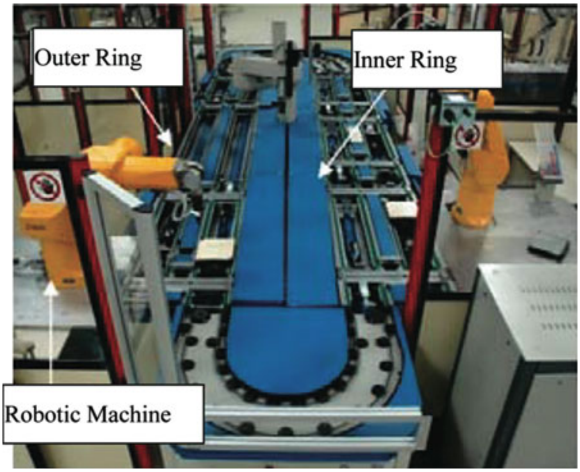


Fig. 42.2 The SISTRE system

working station is. This deviation is performed automatically by reading in certain pre-defined points the RFID tag of the pallet. The working stations are situated on the outer ring. Each station is equipped with pneumatic actuators (pushers, pullers, and indexers) and electric actuators (stoppers) as well as a certain number of RFID reader/writers. Thus, RFID technology allows to identify and locate each pallet as well as to provide the information of the required operation in a certain station. The upper level structure of SISTRE is composed by 5 stations, 2 shifts, a general distribution block, and a security fence. In total the system has 107 important components which need to be facilitated in the ontology model.

42.4 Ontology Implementation

This case study aims at demonstrating several of the capabilities provided by the ontology model. As a starting point we have the SMAC ontology model (Fig. 42.1) and the bill of material (BOM) of SISTRE, which also contains a description for each component/part of the system. The concept is that the SMAC ontology model is utilized for facilitating the information and the data about SISTRE "as it is manufactured" and the data generated during its usage phase. The model is extended according to the usage requirements.

42.4.1 The System Architecture

The basic understanding of the usage and capabilities of the system architecture is a key element for the reader to be able to recreate the presented applications and to use the architecture in future applications. The system architecture we are utilizing is shown in Fig. 42.3. It contains the Protégé OWL-DL editor with the following plug-ins: Data Master, Queries Tab, SWRL, SQWRL, Jess rule engine, and the DL-Reasoner "Pellet".

The functionality of the system is as follows. Firstly, the model is developed in the Protégé-OWL editor as an OWL-DL ontology containing the classes, relationships between classes and attributes. Then, DL rules are added to the classes as restrictions to define them according to requirements. These definitions of each class are machine understandable. In the next step instances are loaded into the model either manually or from spreadsheets through the Data Master. The use of DLs provides the benefit of making the concepts of each class of the model machine understandable. Instances, classes (representing human concepts) and relationships among classes (representing roles of the concepts in real life) are the building blocks used by the ontology to describe the domain. DL rules are understood by the DL-reasoner and support the extensibility of the model. They support the model user with information about inconsistencies, synonyms (equivalencies between classes) and re-classification of the class-hierarchy according to the concepts that each class represents. Finally, the instances containing the data are inferred under the right classes in the model. Therefore, the use of DL rules prevents automatically the creation of data miss-location or of data duplicates. Furthermore, in Fig. 42.3 there is a triangle shaped loop of information flow from OWL-DL, to SWRL, from SWRL and OWL-DL to Jess Rule Engine, and finally back to OWL-DL. This loop describes a process which consists of the following steps: in the first step the SWRL rules are created; in the second step the SWRL rules and the OWL knowledge are transferred to Jess (two input arrows to Jess in Fig. 42.3 one from SWRL and one OWL-DL); in the third step the Jess rule

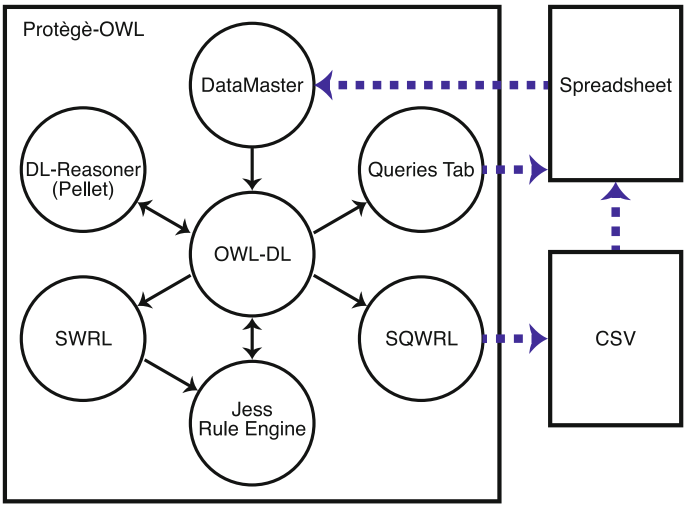


Fig. 42.3 System architecture

engine is executed to infer knowledge (according to SWRL rules and the OWL knowledge it has read); and finally the inferred knowledge from Jess is transferred into the OWL ontology. Since Jess does not read all OWL axioms, Pellet DL-reasoner should be executed at this point to ensure that the OWL model is still consistent. This step by step process is used for inferring instances of selected parts of the OWL model under the right classes and then for returning this knowledge back to OWL. This part specific return of instances is not possible to be performed by the DL-reasoner, since the DL-reasoner firstly, needs to read the whole model and check it for its consistency and then, to continue to perform the inference. The loop is also used when instances and data is introduced using spreadsheets. Different ways for querying the model apart from the above mentioned loop are: the DL-reasoner, the Queries tab and the SQWRL. Each one of these tools is used under different circumstances and it depends on the type and the nature of the query.

42.4.2 Implementation Process

The BOM contains all the 107 components of the system with a description and a separate component code for every distinct component. In total there are 67 distinct components which means that several components are used in multiple places on the system and hence, exist multiple times in the BOM. In the rest of this work, the following naming conventions are used: names of classes are written in boldface and capitalized/lower case Arial (i.e., **Product_MOL**). Names of attributes and relationships are capitalized/lower case Courier New (i.e., `hasParent`) while names of instances are in italics Arial (i.e., *Screw_1*).

The model of Fig. 42.1 is extended to facilitate the new information. Firstly, all data from the BOM has been loaded on the **Physical_Product** class as described in paragraph 4.1. Thus, one new instance per component has been created on the **Physical_Product** class. Several components are complex and they are composed of two or more components. This is expressed by relating each complex instance to other instances of the **Physical_Product** class through the properties `hasParent/isParentOf`. Secondly, the 67 different components of the BOM, were loaded in the developed model as instances of the **Physical_Product_Group** class. The most important attribute of these instances is their component code, which is stored as the `Group_Code`. This code is unique per distinct component and thus, if there are differently named instances of the **Physical_Product_Group** class and they have the same `Group_Code` this means that the instances are logically the same instance. This may be described by SWRL rules and it is well understood by the jess rule engine and hence, duplicates are rejected. An example of a SWRL rule is: Physical_Product_Group(?pgx)^Physical_Product_Group(?pgy)^Group_Code(?pgx,?gcx)^Group_Code(?pgy,?gcy)^differentFrom(?pgx,?pgy)^swrlb: equal(?gcx,?gcy)- >Rejected(?pgy)

This rule reads the instances of **Physical_Product_Group** class in pairs, compares their `Group_Code` and if the `Group_Code` is the same, it moves to a classes named "Rejected" one of the two instances. When this rule is executed by the jess rule engine all the duplicates are moved away from the **Physical_Product_Group** class. In this way if a component exists more than once in the BOL, it is loaded only one time under the **Physical_Product_Group** class. Therefore, the **Physical_Product_Group** class contains one instance per `Group_Code`. Finally, each instance of the Physical_Product class is related to one instance of the **Physical_Product_Group** class.

At this stage all the data of the BOM components is represented as an instance under the **Physical_Product** class, no matter if the component is complex or if it is part of a part of another part. Thus, it is not visible to which part of the actual system each component belongs; information is hidden under the different parts of the model and no proper reasoning may be applied to represent the information. In other words the system architecture of paragraph 4.1 has not been exploited. In the following paragraph we extend the model to facilitate the instances according various criteria and we make an extensive use of DL rules and the reasoner in order to check the model for inconsistencies, equivalencies, infer re-classification of the class hierarchy and infer the instances under the new sub-classes.

42.4.3 Extending the Model Using DL Rules

In order to provide data transparency and to provide a user-friendly solution, the **Physical_Product** class of Fig. 42.1 has been extended with 107 sub-classes, one for each component, which are all placed at the first level of abstraction of the class. Also, an instance *SISTRE_1* has been added under **Physical_Product** class which facilitates the one (or more) SISTRE systems (which in the future will be) in this model. Also, a sub-class named **SISTRE** has been added under the **Physical_Product** class. Each one of these sub-classes is defined by a DL rule which defines which instances should be instances of this class. For example the **SISTRE** class should contain all the instances which have as direct `hasParent` the instance *SISTRE_1* as well as the sub-classes of these instances. The DL rule of the **SISTRE** class is:

$$\mathrm{SISTRE} \equiv \mathrm{Physical\_Product} \cap \ni \mathrm{hasParent}.(\mathit{SISTRE}\_1)$$

This rule means that: a member of **SISTRE** class is a **Physical_Product** whose object property `hasParent` has as value the instance of the **Physical_Product** class, named *SISTRE_1*. The meaning of each sub-class (in total 108 sub-classes) was defined using similar rules. The inheritance of the property `hasParent/isParentOf` for the instance *Virage_Simple_44* is shown in Fig. 42.4. This is an example of the rules that the reasoner follows for re-classifying the class- hierarchy under the **SISTRE** class. The results after the classification are shown in Fig. 42.6. The sub-classes have been re-classified under five levels of abstraction. Also, all

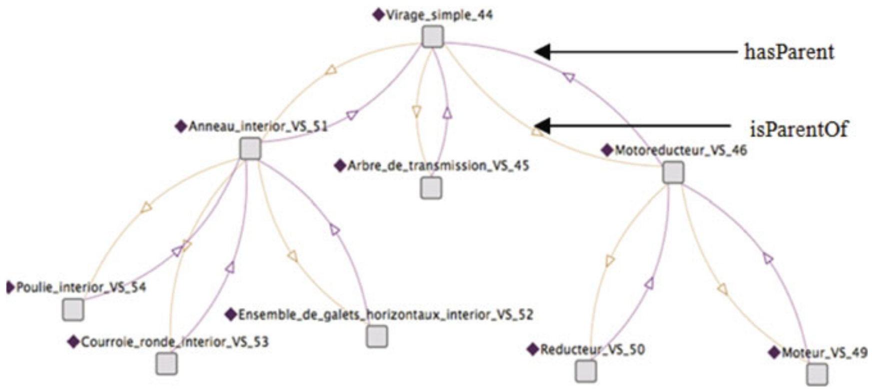


Fig. 42.4 The instance "Virage_Simple_44" and its children

instances which are parts of a specific instance have been inferred under the right class i.e., the parts of *Virage_Simple* are categorized under the **Virage_Simple** class. The DL rules allow the reasoner to understand the content of each class and according the rules to re-classify all the sub-classes at the right level of abstraction at four levels and all instances are categorized under these classes.

To fulfill the requirement of tracking a specific type of component which exists in multiple products, more DL rules where added. This is useful i.e. in cases of discovering that a component is faulty due to a design error and announcing equipment recall. Thus, the manufacturer knows which machines contain this type of component and they are recalled. In order to describe this requirement in the model, new sub-classes were added to the **Physical_Product** class containing a DL rule checking which each instance of **Physical_Product** class is related to a specific instance of the **Physical_Product_Group** class with a specific `Group_Code`. For example, for collecting all the components which belong to "Presence Detector group" we created the **Presence_Detector** class with the following restriction:

$$\text{Presence\_Detector} \equiv \text{Physical\_Product} \cap \ni$$
$$\text{Physical\_Product2Physical\_Product\_Group.}(\textit{Detecteur}-\text{Presence\_Group})$$

This rule means that: a member of **Presence_Detector** is a **Physical_Product** whose object property `Physical_Product2Physical _Product_Group` has as value the instance of the **Physical_Product_Group** class, named *Presence_Detector*. The results of the inference are shown in Fig. 42.5. As long as the model is extended with classes containing DL rules, the reasoner will categorize the new classes under the right position in the class-hierarchy and it will check if the newly created classes actually already exist. An example is shown in Fig. 42.7 in which the class **Class_1** was created having DL rules. The reasoner read the rules, checked them for their consistency, and then, re-classified **Class_1** as equivalent (to be the same as) with **Presence_Detector** class (right column in Fig. 42.7).

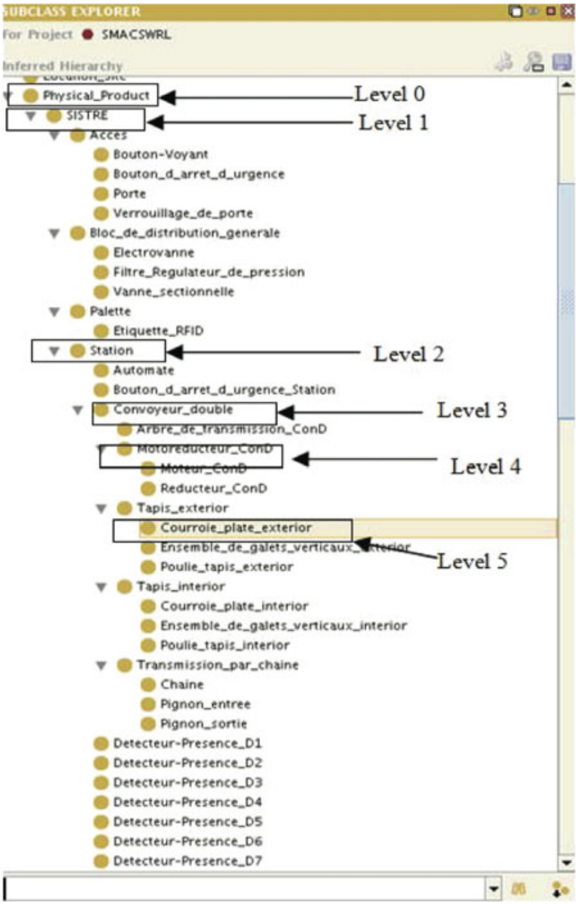


Fig. 42.5 Instances have been inferred under the detect presence class

The reasoner reads the DL rules and: when the inference is fired it reads the properties (Fig. 42.4) infers the instances under the right classes (Fig. 42.5); and when the classification is fired it re-classifies the sub-classes under the right level of abstraction (Fig. 42.6). Also, in the case that one or more classes have different names but they contain logically the same rule are recognized as equivalent (Fig. 42.7). In this way whenever new instances or classes are added in the model, the instances will be categorized under the right classes and the classes under the right level of abstraction. These capabilities are very important in cases of merging one or more ontology models shared among partners. Taking for granted that all partners started from the same initial SMAC-Model and they extended their models using DL rules according to their needs, when merging them, the reasoner will figure out the new elements of each model. The merged model will contain only the new elements avoiding duplicates. Thus, data integration and interoperability between different model variations are preserved.

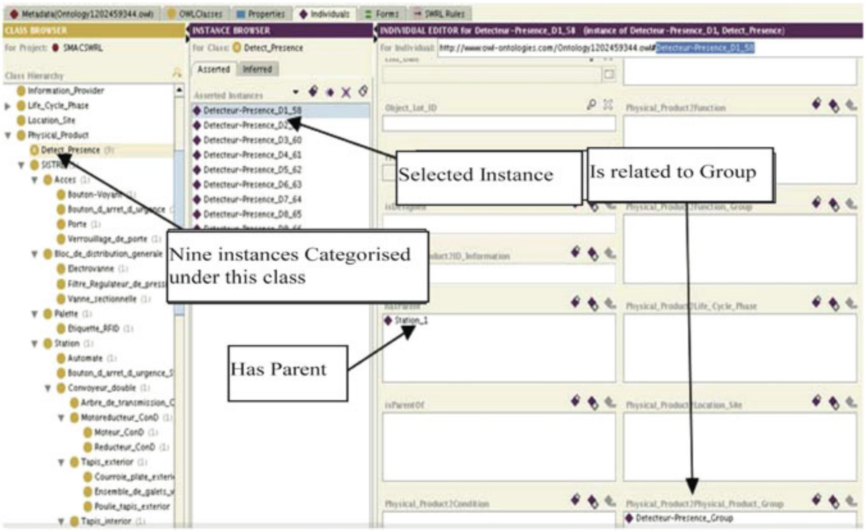


Fig. 42.6 Class hierarchy under the SISTRE class in five levels of abstraction

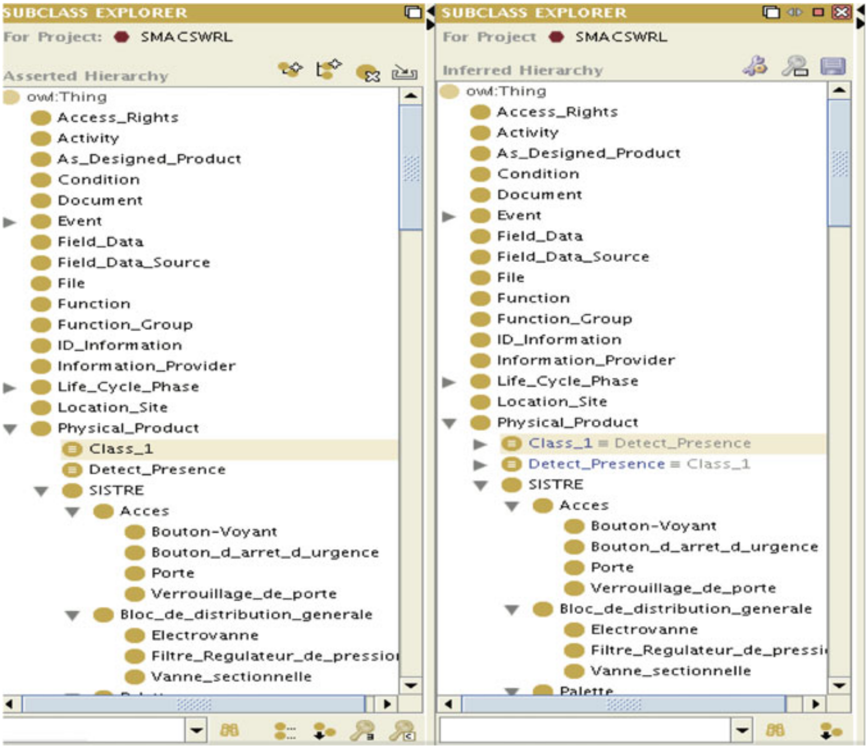


Fig. 42.7 The equivalencies as they are recognized by the reasoner

42.5 Discussion and Conclusions

This work describes an application of the SMAC-Model and shows how it is generic, extensible and it supports data integration. All the components of SISTRE and their information have been loaded on the model. Then, the model has been extended with sub-classes using DL rules. When, the reasoner is executed, the model is checked for its consistency; classes for equivalencies and they are re-classified to the right position in the model; added and/or existing instances will be categorized under the right classes. These capabilities also provide support in cases of merging one or more ontology models by preserving data integration and interoperability among the models. Future work includes the development of a demonstrator with sensors and the collection of field data. Then, web services will be used for accessing the data and advanced techniques to treat it efficiently. In this part excessive use of SWRL and Jess rules is expected. A challenge remains the handling of events-alarms and the accuracy of the system in predicting maintenance activities.

Acknowledgments This work was carried out in the framework of SMAC project (Semantic-maintenance and life cycle), supported by Interreg IV program between France and Switzerland.

References

1. Matsokis A, Karray MH, Morello-Chebel B, Kiritsis D (2010) An ontology-based model for providing semantic maintenance. Adv Maintenance Eng 1(1):1250–1260
2. Baader F, Calvanese D, McGuinness D, Nardi D, Patel-Schneider PF (2003) The description logic handbook: theory, implementation, and applications. Cambridge University Press, Cambridge
3. O'Connor M (2007) SWRL protégé plug-in. Available from: http://protegewiki.stanford.edu/index.php/SWRLTab. Cited May 2011
4. Sandia National Laboratories (2011) The Jess rule engine. Available from: http://herzberg.ca.sandia.gov/. Cited May 2011
5. Muller A, Crespo A, Iung B (2008) On the concept of e-maintenance: review and current research. J Relib Eng Syst Saf 93:1165–1187
6. MIMOSA (1998) MIMOSA (Machinery information management open systems alliance) Available from: www.mimosa.org. Cited May 2011
7. Mathew A, Zhang L, Zhang S, Ma L (2006) A review of the MIMOSA OSA-EAI database for condition monitoring systems. In: Proceedings of the 1st world congress on engineering asset management (WCEAM), Gold Coast, Australia. Springer, London, pp 837–846
8. Voisin A, Levrat E, Cocheteux P, Iung B (2010) Generic prognosis model for proactive maintenance decision support-application to pre-industrial e-maintenance test bed. J Intell Manuf 21(2):177–193
9. Matsokis A, Zamofing S, Kiritsis D (2010) Ontology-based modelling for complex industrial asset lifecycle management: a case study. In: Proceedings of the 7th international conference on product lifecycle management (PLM'10), Bremen, Germany (in press)

Chapter 43
A Mathematical Formulation of the Problem of Optimization of Inspection Planning in Asset Management

J. Mashford, D. Marlow, D. Marney and S. Burn

Abstract The problem of optimal planning for inspection of assets in engineering asset management is formulated in the case that the inspections take place at the beginning of the planning period. It is shown that the problem has the form of a 0-1 knapsack problem. The formulation is applicable to systems consisting of multiple assets. The concept of risk, as usually defined in asset management, arises from the formulation in a natural way and plays an important part in identifying optimal inspection plans. Two special cases are considered. In the first special case, where the inspection budget is large, the optimal inspection plan involves inspecting all assts whose risk is sufficiently large. In the second special case, where the costs and expected benefits associated with each asset are the same and the conditional probability of failure of each asset given that it has been inspected is negligible, the optimal inspection plan involves prioritization of assets on the basis of risk alone.

43.1 Introduction

Inspection and subsequent possible maintenance of engineering assets (repairable assets may be repaired while non-repairable assets may be replaced) is carried out subject to budgetary constraints and the level of such activities influences the total cost of operating the asset or collection of assets. For low levels of service the total cost associated with the operation of a collection of assets will decrease as the level of service increases while for high levels of service it will increase as the level of service increases. At some point in between these extremes there is an optimal level of service for which the total cost is least [1].

J. Mashford (✉) · D. Marlow · D. Marney · S. Burn
Commonwealth Scientific and Industrial Research Organization, PO Box 56 Highett, VIC 3190, Australia
e-mail: john.mashford@csiro.au

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_43,

Optimization of inspection would result in a plan to decide which assets to inspect and, possibly, when they should be inspected. There is a distinction between proactive assets for which inspection or maintenance is scheduled prior to failure and reactive assets which are allowed to run to failure (though inspection may occur for reactive assets once failure has occurred). Inspection planning may be carried out by prioritizing assets for inspection and selecting assets to inspect starting from the highest priority assets and proceeding to lower priority assets until the inspection budget is exhausted.

The problem of optimizing the inspection of systems of multiple assets is complex and analytical solutions have only been obtained for systems consisting of a single asset. For a single asset the problem arises of determining the optimal time interval between inspections. If the interval is small then there will be a smaller cost due to failure occurrences but higher cost due to inspections and preventive repairs or replacements. On the other hand if the interval is large then the costs due to inspection, repair, or preventive replacement will be smaller but the costs due to failure occurrence will be higher. At some point in between there is an optimum value for the interval between inspections. Barlow [2] gives the classical solution to the problem of finding this optimum value in the case of an infinite planning horizon. Legat et al. [3] present a method for obtaining an approximate solution to this problem in the case of a finite time horizon. Ghosh and Roy [4] discuss the optimization of the maintenance of a single component in a process plant on the basis of maximizing the benefit to cost ratio.

Inspection prioritization is frequently carried out on the basis of risk which is a quantity associated with each asset. Risk, in asset management, is usually defined as the product $R = P \times C$ of the probability of failure of the asset times its consequence of failure [1, 5–10]. However, no real justification is given in the literature as to why risk defined as the product in this way should be important in inspection prioritization; why, for example, could not some other function of P and C be the most appropriate quantity to use for prioritization?

The risk-based inspection and maintenance strategy developed by the American Society of Mechanical Engineers [11] was used as a basis for developing a "base resource document on risk-based inspection" by the American Petroleum Institute [12]. Risk based approaches have been used for the maintenance of medical devices [13]. Khan and Haddara [8] describe a risk-based maintenance methodology in which the quantitative value of risk is used to prioritize inspection and maintenance activities. The level of risk is used to determine the type, means and timing of preventive tasks. Lee et al. [9] describe a method for prioritizing inspection of components of floating, production, storage and off-loading (FPSO) shipping vessels in which priority is based on risk. Khan et al. [14] discuss risk-based inspection modeling of process components and systems including stochastic degradation models and an approach to optimal inspection interval calculation.

Rogers and Grigg [15] describe a model for prioritization of water pipe renewal including a risk module providing the capability for prioritizing pipe renewal activities based on risk management. Burn et al. [1] describe an overall asset management strategy for water pipes using a risk-based approach.

Ugarelli and Federico [16] developed an optimization model for the management of wastewater pipeline networks in which the total cost associated with an asset during any year was set equal to the sum of the expected direct cost of maintenance and the risk of failure. Willcocks and Bai [10] discuss risk-based inspection and integrity management (RBIM) of pipeline systems which is a means of focusing and optimizing the use of resources to 'high risk areas'.

In the present paper we consider the optimization of the inspection plan for a collection of assets over some finite planning horizon where it is assumed that the inspections take place at the beginning of the planning period under consideration. An inspection plan is represented by a collection $\{x_i\}$ of 0-1 decision variables x_i which indicate whether asset i is to be inspected. One might consider that in order to generalize this to the case of inspections being carried out throughout the planning period one can introduce time-dependent decision variables x_{it}. However, there is a fundamental problem with evaluating fixed inspection plans of this type. The problem is due to our incomplete knowledge about the behavior of assets and the fact that this knowledge can only be represented in probabilistic terms. Thus suppose, for example, that we have a fixed inspection plan $\{x_{it}\}$ which specifies that asset i should be inspected in December 2011. To evaluate this inspection plan we would need to consider all possible histories of the evolution of asset i and compute the expected net benefit over all histories. However, some of the possible histories would involve a failure in November 2011. In such cases, unplanned repairs would be carried out in November 2011 and any inspection planned for December 2011 would ideally be cancelled or rescheduled. Thus some histories of asset i invalidate the inspection plan $\{x_{it}\}$. This must be taken into account in the mathematical formulation of the optimization problem. In fact, the objects being optimized must be *inspection policies* rather than fixed inspection plans. An inspection policy may give tentative inspection times for each asset but must also specify a policy for cancellation or rescheduling of planned inspections if it is appropriate to do so.

This problem is taken into consideration in the work of Shirmohammadi et al. [17] and [18] where they consider the case of a single asset (machine) and optimize policies based on a decision variable for the time interval between maintenance and another decision variable specifying a cutoff time for skipping maintenance. Another approach which takes into consideration the dependency of an inspection event on the recent failure history of an asset is the use of decision trees [19, 20]. This approach becomes unwieldy in the case of more than one asset.

The problem is circumvented in the present paper by considering inspections to take place at the beginning of the planning period. In reality assets are inspected at various times throughout the planning period. The techniques described in this paper can be used to determine which assets should be inspected during a planning period. Then practical considerations, such as available resources, can be used to determine the exact inspection schedule. The main contribution of this paper is the formulation of the inspection optimization problem in the case of systems of multiple assets and the proof that the concept of risk defined in the usual way is relevant and important to the solution of this problem. In Sect. 43.2, the inspection optimization problem is formulated and is shown to have the form of a 0-1 knapsack problem. In Sect. 43.3, two special cases are considered including one leading to the conventional risk-based prioritization and a conclusion to the paper is given in Sect. 43.4.

43.2 Inspection Strategy Optimization

For various assets, data can be collected so that the expected characteristics of failure can be inferred statistically. Consider a collection of n assets over some period of time for which we know, or at least have a reasonable estimate of,

p_i = the probability of failure of asset i
q_i = the probability of failure of asset i given that it has been inspected and repaired if need be after the inspection
c_i = the consequence of failure of asset i
d_i = the cost of inspecting asset i
e_i = the expected cost of repairing asset i
b_i = the expected benefit of repairing asset i,
for each i = 1, …, n.

It is assumed that the consequence c_i is measured in units of cost (e.g. $) so that, for example, a triple bottom line [21] consequence can be converted to an equivalent $ amount by appropriately weighting each of the three consequence results. The cost e_i is the expected cost of repairing asset i once inspected if any repairs are deemed necessary. The expected benefit b_i arises as a result of future cost savings associated with the fact that asset i will not need to be inspected for an extended period of time and it is less likely to fail at future times.

Now suppose that we want to evaluate an inspection strategy specified by 0-1 variables x_i where

$$x_i = \begin{cases} 1 & \text{if asset number i is inspected} \\ 0 & \text{otherwise} \end{cases}$$

Since we only have probabilities or expected values of some of the variables describing the system we consider a sequence of trials where for each j the condition of the system in trial j is specified by the following.

$$f_i^{(j)} = \begin{cases} 1 & \text{if asset i fails in trial j} \\ 0 & \text{otherwise} \end{cases}$$

$e_i^{(j)}$ = the cost of repairing asset i in trial j if it is inspected (if no repair or replacement is deemed necessary this cost is 0)
$b_i^{(j)}$ = the benefit of repairing asset i in trial j

Then
Total cost of adopting inspection strategy {xi} for trial j

$$\begin{aligned}&= \text{Cost of inspecting the assets which are inspected}\\&\quad + \text{Cost of repairing or replacing the assets which are inspected}\\&\quad + \text{Cost of failure of the assets which fail}\\&= \sum\nolimits_{i=1}^{n} d_i x_i + \sum\nolimits_{i=1}^{n} e_i^{(j)} x_i + \sum\nolimits_{i=1}^{n} c_i f_i^{(j)}\end{aligned} \tag{43.1}$$

Also,
Benefit from adopting inspection strategy {xi} for trial j

$$= \sum\nolimits_{i=1}^{n} b_i^{(j)} x_i. \tag{43.2}$$

Therefore
Net benefit of adopting inspection strategy $\{x_i\}$ for trial j

$$= \sum\nolimits_{i=1}^{n} \left(\left(b_i^{(j)} - d_i - e_i^{(j)} \right) x_i - c_i f_i^{(j)} \right). \tag{43.3}$$

Now:

$$\begin{aligned}&\text{Expected net benefit of adopting inspection strategy } \{x_i\}\\&\quad = \text{limit } (m \to \infty) \text{ of the average net benefit of adopting strategy } \{x_i\} \text{ over m trials}\\&\quad = \lim_{m\to\infty} \left(\frac{1}{m} \sum\nolimits_{j=1}^{m} \left[\sum\nolimits_{i=1}^{n} \left(\left(b_i^{(j)} - d_i - e_i^{(j)} \right) x_i - c_i f_i^{(j)} \right) \right] \right)\\&\quad = \sum\nolimits_{i=1}^{n} \left(\left(\lim_{m\to\infty} \frac{1}{m} \sum\nolimits_{j=1}^{m} b_i^{(j)} - d_i - \lim_{m\to\infty} \frac{1}{m} \sum\nolimits_{j=1}^{m} e_i^{(j)} \right) x_i - c_i \lim_{m\to\infty} \frac{1}{m} \sum\nolimits_{j=1}^{m} f_i^{(j)} \right)\\&\quad = \sum\nolimits_{i=1}^{n} \left((b_i - d_i - e_i) x_i - c_i \lim_{m\to\infty} \frac{1}{m} \sum\nolimits_{j=1}^{m} f_i^{(j)} \right)\end{aligned} \tag{43.4}$$

If an asset is not inspected then it will fail with long run frequency equal to its probability of failure. Therefore

$$x_i = 0 \Rightarrow \lim_{m\to\infty} \frac{1}{m} \sum\nolimits_{j=1}^{m} f_i^{(j)} = p_i, \ \forall i = 1, \ldots, n. \tag{43.5}$$

On the other hand, if an asset is inspected then it will fail with long run frequency equal to its conditional probability of failure given that it has been inspected. Therefore

$$x_i = 1 \Rightarrow \lim_{m\to\infty} \frac{1}{m} \sum\nolimits_{j=1}^{m} f_i^{(j)} = q_i, \ \forall i = 1, \ldots, n. \tag{43.6}$$

It follows that

$$\lim_{m\to\infty} \frac{1}{m} \sum_{j-1}^{m} f_i^{(j)} = p_i(1 - x_i) + q_i x_i,\ \forall i = 1, \ldots, n. \tag{43.7}$$

Thus

$$\begin{aligned} &\text{Thus Expected net benefit of adopting inspection strategy } \{\mathrm{x_i}\} \\ &= \sum_{i=1}^{n} ((b_i - d_i - e_i)x_i - c_i p_i(1 - x_i) - c_i q_i x_i) \\ &= \sum_{i=1}^{n} ((b_i - d_i - e_i + c_i p_i - c_i q_i)x_i - c_i p_{ii}) \\ &= \sum_{i=1}^{n} ((b_i - d_i - e_i + r_i - s_i)x_i - r_i) \\ &= \sum_{i=1}^{n} ((b_i - d_i - e_i + r_i - s_i)x_i) - \sum_{i=1}^{n} r_i \end{aligned}, \tag{43.8}$$

where

$$\begin{aligned} r_i &= p_i c_i \\ s_i &= q_i c_i. \end{aligned} \tag{43.9}$$

$\mathrm{r_i}$ is called the risk of failure of asset i and is the product of the assets probability of failure and its consequence of failure. $\mathrm{s_i}$ is the risk of failure of asset i given that it has been inspected.

The cost of carrying out inspections and associated repairs, not counting repairs associated with failures, is

$$\cos t = \sum_{i=1}^{n} (d_i + e_i)x_i. \tag{43.10}$$

Suppose that we have an inspection budget B and we are seeking an inspection strategy $\{\mathrm{x_i}\}$ which maximizes the expected net benefit such that the inspection cost does not exceed the inspection budget. This is equivalent to solving the following optimization problem.

$$\max \sum_{i=1}^{n} (b_i - d_i - e_i + r_i - s_i)x_i, \tag{43.11}$$

such that $\sum_{i=1}^{n} (d_i + e_i)x_i \le B$.

Note that the second term in Eq. (43.8) is constant (independent of $\{\mathrm{x_i}\}$) and so can be omitted from the maximization. This is in the form of the 0-1 knapsack problem which is the well studied problem of optimally choosing from a number of objects of various sizes and values which ones to carry in a knapsack of a fixed size [22]. The 0-1 knapsack problem can be solved in pseudo-polynomial time using dynamic programming and can be solved approximately in polynomial time using an approximation scheme which uses the pseudo-polynomial time algorithm as a subroutine [22].

43.3 Two Special Cases

As a special case suppose that the inspection budget is infinite, or at least larger than $\sum_{i=1}^{n}(d_i + e_i)$. Then the constraint of the optimization problem is satisfied for all inspection strategies $\{x_i\}$. In this case the optimal solution is obtained by inspecting all assets for which the risk satisfies

$$r_i \geq d_i + e_i - b_i + s_i. \tag{43.12}$$

As another special case suppose that $d_i + e_i$ is the same (or approximately the same) for each asset i, with

$$d_i + e_i = g, \forall i = 1, \ldots, n, \tag{43.13}$$

that b_i is the same (or approximately the same) for each i, with

$$b_i = b, \forall i = 1, \ldots, n, \tag{43.14}$$

and $b \geq g$, and that the probabilities q_i are negligible.

Then the optimization problem to be solved is

$$\max \sum_{i=1}^{n} (b - g + r_i)x_i, \tag{43.15}$$

such that

$$\sum_{i=1}^{n} x_i \leq {}^{B}\!/_{g}$$

The optimal solution to this problem is obtained by listing all the assets in order of risk and choosing the top k of them to inspect, where

$$\begin{aligned} \mathrm{k} &= [\mathrm{B}/\mathrm{g}] \\ &= \text{greatest integer less than or equal to } \mathrm{B}/\mathrm{g}. \end{aligned}$$

The optimal inspection strategy in this special case coincides with the strategy dictated by risk-based asset management. When it is not the case that the inspection costs and benefits are constant (with cost not exceeding benefit) and the conditional probabilities of failure are negligible, the optimal inspection strategy does not coincide with the risk-based asset management approach but is obtained by solving the 0-1 knapsack problem (43.11). The solution to this problem will indicate which assets should be inspected during the planning period. Then practical considerations such as availability of resources can be used to determine the order and timing of the inspections.

43.4 Conclusion

The problem of optimal planning for inspection of assets in engineering asset management is complex. To take into account the time at which inspections are made it would be necessary to consider deterioration of assets over time using concepts from reliability theory. Also, it would be necessary to optimize over inspection policies rather than inspection plans and concepts from decision theory would be required. Such investigations have, in the literature to date, only been carried out for the case of systems consisting of a single asset.

In the present paper the problem of planning for inspection of assets in systems consisting of multiple assets is formulated in the case that the inspections take place at the beginning of the planning period. It is shown that the problem has the form of a 0-1 knapsack problem. The concept of risk, as usually defined in asset management, arises from the formulation in a natural way and plays an important part in identifying optimal inspection plans. Two special cases are considered. In the first special case, where the inspection budget is large, the optimal inspection plan involves inspecting all assets whose risk is sufficiently large. In the second special case, where the costs and expected benefits associated with each asset are the same and the conditional probability of failure of each asset given that it has been inspected is negligible, the optimal inspection plan involves prioritization of assets on the basis of risk alone.

Further research will involve the consideration of systems of multiple assets in which inspections can take place at any time over the planning period.

The work described in this paper has the potential to be used for the development of practical systems for inspection planning if probabilities of failure can be estimated by statistically analyzing failure data or by other means, consequences of failure can be estimated and the required costs and expected benefits can be estimated. Such estimates may be made with respect to covariates such as asset type and asset properties. Once these quantities have been estimated for a collection of assets the optimization problem (43.11) can be solved using appropriate software to determine which assets should be inspected after which the order and timing of inspections can be determined on the basis of practical considerations by the asset manager.

Acknowledgments The authors would like to thank David Beale and Mike Rahilly for very helpful discussions.

References

1. Burn S, Marlow D, Moglia M, Buckland P (2007) Asset management for water infrastructure. Water Asset Manage Int 3(2):12–18
2. Barlow RE, Proschan F (1965) Mathematical theory of reliability. Wiley, New York

3. Legat V, Zaludova AH, Cervenka V, Jurca V (1996) Contribution to optimization of preventive replacement. Reliab Eng Syst Saf 51(3):259–266
4. Ghosh D, Roy S (2009) Maintenance optimization using probabilistic cost-benefit analysis. J Loss Prev Process Ind 22(4):403–407
5. Kaplan S, Garrick BJ (1981) On the quantitative definition of risk. Risk Anal 1(1):11–27
6. IPWEA (2006) International infrastructure management manual, 3rd edn. Australia/NZ Edition. Association of Local Government Engineering NZ Inc.
7. Faber MH, Stewart MG (2003) Risk assessment for civil engineering facilities: critical overview and discussion. Reliab Eng Saf Syst 80(2):173–184
8. Khan FI, Haddara MM (2003) Risk-based maintenance (RBM): a quantitative approach for maintenance/inspection scheduling and planning. J Loss Prev Process Ind 16(6):561–573
9. Lee AK, Serratella C, Wang G, Basu R, Spong R (2007) Multilevel risk-based inspection scheme for FPSOs. Mar Technol 44(2):118–124
10. Willcocks J, Bai Y (2000) Risk based inspection and integrity management of pipeline systems. In: Proceedings of the 10th international offshore and polar engineering conference, vol 2 pp 285–294
11. ASME (1991) Research task force on risk based inspection guidelines, risk based inspection development of guidelines: general document. American Society of Mechanical Engineers, Washington, DC
12. API (1995) Base resource document on risk based inspection for API committee on refinery equipment. American Petroleum Institute, Washington DC
13. Ridgeway M (2001) Classifying medical devices according to their maintenance sensitivity: a practical risk-based approach to PM program management. Biomed Instrum Technol 35(3):167–176
14. Khan FI, Haddara MM, Bhattacharya SK (2006) Risk-based integrity and inspection modeling (RBIIM) of process components/systems. Risk Anal 26(1):203–221
15. Rogers PD, Grigg NS (2009) Failure assessment modeling to prioritize water pipe renewal: two case studies. J Infrastruct Syst 15(3):162–171
16. Ugarelli R, Di Federico V (2010) Optimal scheduling of replacement and rehabilitation in wastewater pipeline networks. J Water Resour Plann Manage 136(3):348–356
17. Shirmohammadi AH, Love CE, Zhang ZG (2003) An optimal maintenance policy for skipping imminent preventive maintenance for systems experiencing random failures. J Oper Res Soc 54(1):40–47
18. Shirmohammadi AH, Zhang ZG, Love CE (2007) A computational model for determining the optimal preventive maintenance policy with random breakdowns and imperfect repairs. IEEE Trans Reliab 56(2):332–339
19. Straub D, Faber M (2005) Risk based inspection planning for structural systems. Struct Saf 27(4):335–355
20. Straub D, Faber M (2006) Computational aspects of risk-based inspection planning. Comput-Aided Civ Infrastruct Eng 21(3):179–192
21. Brown D, Dillard J, Marshall RS (2006) Triple bottom line: a business metaphor for a social construct. Portland State University, School of Business Administration, Working paper, Portland, OR
22. Kellerer H, Pferschy U, Pisinger D (2004) Knapsack problems. Springer, New York

Chapter 44
Self-Management Process in S-Maintenance Platform

M. H. Karray, B. C. Morello, C. Lang and N. Zerhouni

Abstract E-maintenance systems are considered as platforms integrating various systems in the maintenance scope, but these platforms provide only the services provided by their integrated systems. S-maintenance platform is built in the aim to provide dynamic services thanks to its core components, especially its knowledge base. This paper focus the exploitation of the s-maintenance architecture's components to define two new processes that we called self management and self learning processes. These processes allow the automatic acquirement and integration of knowledge in the knowledge base and the dynamic evolution of the platform behavior.

44.1 Introduction

The role of maintenance is strategic in the industrial environment. Hence, while having this interest in company, industry has been seeing the developments of different generations of maintenance systems. New technologies of information and communication technologies (ICT) have helped to establish and evolve these systems. Thanks to ICT, Web emergency and Internet, the achievement of

M. H. Karray · B. C. Morello · C. Lang (✉) · N. Zerhouni
Automatic Control and Micro-Mechatronic Systems Department, 24 Rue Alain Savary
25000 Besançon, France
e-mail: christophe.lang@univ-fcomte.fr

M. H. Karray
e-mail: hedi.karray@femto-st.fr

B. C. Morello
e-mail: brigitte.morello@femto-st.fr

N. Zerhouni
e-mail: noureddine.zerhouni@femto-st.fr

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_44,

maintenance services and monitoring can be performed automatically, remotely and through various distributed information systems. Hence the emergence of the concept of services offered through maintenance architectures, ranging from autonomic systems to integrated systems where cooperation and collaboration are vital to any operation. E-maintenance concept that is widespread today in the industry refers to the integration of the information and communication technologies (ICT) within the maintenance strategy and/or plan [1] to face the new needs emerging from innovate ways for supporting production (e-manufacturing), business (e-business). But services provided by the existent e-maintenance systems are not dynamic and not flexible. In fact, services provided by these systems are predefined and cannot adjust and evolve their behavior. They aren't implanted to be able to manipulate new generated knowledge and to operate in concordance with this new knowledge and as well as new situations confronted by the maintenance system.

Consequently, in previous work we proposed a new concept presenting the tomorrow's generation of maintenance system that we called s-maintenance ("s" like semantic). It is about a knowledge oriented system allowing the emergence of dynamic services of maintenance thanks to knowledge engineering capabilities. The core of the maintenance system is a knowledge based system which extracts knowledge and generates new knowledge thanks to the inference engine, as well as having self-learning and self-management capabilities. This system is based on expert knowledge of maintenance domain shared between all components of the maintenance system [2]. So, having the Knowledge that the concept of s-maintenance should be realized, we proposed a first possible architecture of a computing platform of s-maintenance based on knowledge formalized by domain ontology [3].

In this work, the architecture of s-maintenance system is exploited to define two new processes that we called self management and self learning processes allowing the automatic acquisition and integration of knowledge in the knowledge base of the platform. Components of this architecture exploit this knowledge base especially the ontology part about maintenance processes in order to perform a dynamic evolution of the platform behavior. These new processes enable the s-maintenance platform to attain some level of intelligence allowing it to self-manage some tasks in the maintenance process.

The self learning process is ensured by a specific component called "Traceability Manager". This latter saves the platform's interaction in the knowledge base and then does some operations to extract the needed knowledge to be learned, after that it applies the machine learning algorithm C4.5 to obtain new knowledge rules. After the validation of the learning results, the self-management process adapts them to the knowledge base structure in order to update this latter and then self-manages the updated process. The rest of this paper will be organized as follow. Section 44.2 is devoted to remember the architecture of the s-maintenance platform. Section 44.3 is devoted to define and explain these two processes and to present their integration architecture in the proposed platform. A motivating example with the explanation of the used part of the maintenance ontology will be presented in Sect. 44.4. Sections 44.5 and 44.6 will focus on the functioning of the self-learning and self-management processes in the s-maintenance platform. Finally, future works and some conclusions will be presented in Sect. 44.7.

44.2 S-Maintenance Platform

44.2.1 Overview

E-maintenance is defined [2] as the carrying out of maintenance where all technical, administrative, and managerial actions or activities interact and collaborate electronically, using network or telecommunications technologies. Regarding the e-maintenance system, we define it as a collaboration system offering a set of software components and software services of maintenance support (integrated and/or distant) all that enable maintenance actors to find each other and the information they need and to be able to communicate and work together to achieve maintenance process via a variety of devices.

Vis-à-vis, s-maintenance, it comes as an evolution in maintenance systems extending the concept of e-maintenance and transforms the maintenance system from an integration tool to the core of the maintenance process.

s-maintenance concept is defined [2] as the carrying out of maintenance based on the domain expert knowledge, where systems in the network manage this knowledge (formalization, acquisition, discovery, elicitation, reasoning, maintenance use and reuse) and share the semantics to emerge new generation of maintenance services (as self-X services, Worker self-management, improvement of various types of maintenance, new collaboration methods, etc.) while including e-maintenance characteristics. When we speak about characteristics we mean what distinguish this concept according to its definition.

In the other hand, s-maintenance system is a based knowledge distributed and collaborative system providing self-management services [4] to the overall maintenance process. It is a self learning system [5] having the possibility to evolve its intelligence degree, including computational intelligence and knowledge [6]. Based on semantics of the domain experts knowledge, it is a scalable, adaptable and contextual system (semantic interoperability, user oriented, intelligent HMI, processes orchestration, tracking users interactions, new process generation).

This concept emerges from the aggregation of the powerful knowledge engineering and the integration of new aspects into the maintenance scope. The strong point of this concept is that the knowledge management system presents the core of the information system and allows manipulating the expertise knowledge of the domain.

44.2.2 S-Maintenance Platform Architecture

In previous work, a first possible architecture of a s-maintenance system was proposed. This architecture is a component based system (CBS) respecting the characteristics of s-maintenance [3]. Components of this architecture are organized on two categories, generic components and specific components.

Generic components are components which are typically recommended to be integrated in an intelligence maintenance system which are data acquisition system (SCADA, Supervisory Control And Data Acquisition), Diagnosis system, CMMS (Maintenance Management Computer Aided), an ERP (Enterprise Resource Planning), a documentation system (eDoc), a Prognosis system, a scheduling system and finally a geo-location system. Each system relies on the model of the enterprise, the physical system or the equipment. All of these components can be integrated as extern applications in the system.

The specific components respond to the characteristics of s-maintenance as self-learning and self-management. These components are presented in Fig. 44.1 extracted from [3]. In this work we will only focus on components that are implicated to carry out self-learning and self-management processes.

Coordinator: it coordinates between all the system's components and the integrated applications. In the system core, all components are connected via the coordinator. In addition, this component is also a conduit of communication between users with the system and between applications with the core.

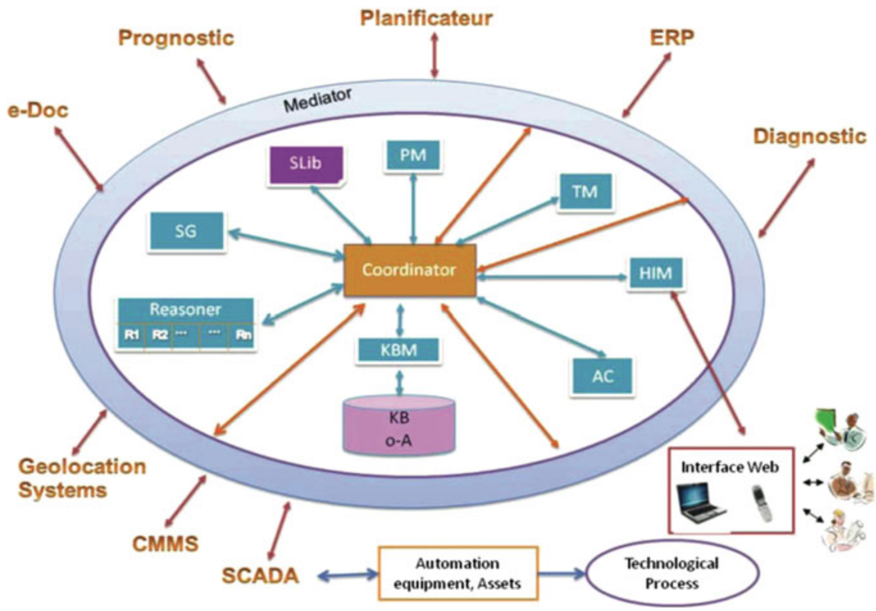


Fig. 44.1 Architecture of s-maintenance platform

KB (Knowledge Base): In general, the ontology and all individual instances of concepts provide a knowledge base. In our case, knowledge base contains the ontology (concepts and rules [ontological meta-model]), ontology (instances of ontological meta-model), data (A: assertions [instances of ontologies]).

Reasoner (Reasoning Engine): The overall goal of a reasoning engine is to seek information and relationships of the knowledge base and provide answers, suggestions and predictions similar to the way a human expert. That is to say, new knowledge is generated from existing knowledge through reasoning engine. The reasoning engine contains general algorithms that can afford to manipulate the knowledge stored in the knowledge base.

PM–Process Manager: It orchestrates all the processes that pass through the system. That is to say that the PM component is responsible for managing the order process passing through the platform.

TM (Traceability Manager): It allows the tracking of all actions done through the system. While all internal communications between the platform and the applications pass through the Coordinator, therefore the TM need just to be connected to the coordinator to trace the overall platform's interactions. The TM allows the tracking of all the interactions between applications, users with the platform and the interactions between components of the platform.

44.3 Self-Learning and Self Management in S-Maintenance Platform

44.3.1 Self-Learning Process

According to Nilsson, learning, like intelligence, covers such a broad range of processes that it is difficult to define precisely. A dictionary definition includes phrases such as "to gain knowledge, or understanding of, or skill in, by study, instruction, or experience," and "modification of a behavioral tendency by experience" [7]. Machine learning refers to a system capable of acquiring and integrating the knowledge automatically. The capability of the systems to learn from experience, training, analytical observation, and other means, results in a system that can continuously self-improve and thereby exhibit efficiency and effectiveness. A machine learning system usually starts with some knowledge and a corresponding knowledge organization so that it can interpret, analyze, and test the knowledge acquired [8]. Make machines learn automatically without human interventions, that is, make machines capable of self-learning [9]. So, while machine learning techniques are considered as the core of any self-learning process, a self-learning system is characterized by the self-adjustment of its behaviors according to this learned knowledge [10].

As mentioned, above the s-maintenance system is a self-learning system having the possibility to evolve its intelligence degree, including computational intelligence and knowledge. It supports the skills based maintenance, focusing skills of knowing how to manage maintenance is a must for the system and is the key to acquiring new skills and sharpening the ability to reason through problems and to surmount challenges. It opens the door to all other learning and facilitates the acquisition of other skills in maintenance management. Skill acquisition is the ability to develop a measurable task repeatedly. The Most important skill learned by the s-maintenance system is the problem solving and decision making skills.

Based on learned skills via the managers, experts and operators interactions' tracking, the s-maintenance platform is able to understand relationships of process and associated data. It will enhance some human trait or experience; obsolesce some other way of process execution, retrieve some old method or experience that was abandoned long ago, and ultimately reverse to create an effect opposite to that originally intended.

According to the definition of the platform, it is a self-learning system which has the ability to upgrade its level of intelligence; it evolves according to the learning from the analysis of actions drawn. To respect this definition, the s-maintenance platform includes the component TM (Traceability Manager) that enables to track and to save all actions done through the platform in the knowledge base as new instances of the ontological model presented above.

44.3.2 Self-Management Process

Computer systems shall manage their own operation without human intervention by self-Management process. Self-Management technologies are expected to pervade the next generation of network management systems [11]. Self-management requires systems to know when and where abnormal behavior occurs, analyze the problem situation, make management plans, and suggest various solutions to the system administrator or manage themselves. This knowledge needs to be made explicit. Self-management knowledge is used to refer to (1) the knowledge that a system has about its internal structure and its dynamic behavior (object-level concepts), (2) knowledge rules to reason about these object-level concepts, and (3) the meta-knowledge to reason about these rules and concepts [12].

From the self-management aspect emerges the aspect of self-management of the maintenance process. It is the capacity to the maintenance platform to make decisions to manage the maintenance process (e.g. planning, preparation and execution) under a trust level for a specific task. Based on the platform acquired knowledge and intelligence level obtained by the platform self-learning process the platform self-manages some tasks (activities) in the maintenance process. Another aspect can be integrated in this scope which is the Worker self-management. This latter is another type of self management. It is a form of workplace decision-making in which the workers themselves agree on choices (for issues like customer care, general production methods, scheduling, division of labor etc.) instead of an owner or traditional supervisor telling workers what to do, how to do it, and where to do it [13].

The self-management of the maintenance process in the s-maintenance platform is ensured by the PM component as mentioned above. This component takes advantages from the ontology of maintenance especially the process view allowing the understanding of the platform behaviors as well as the results of the learning process.

44.3.3 Self-Learning and Self-Management Architecture

During the execution of s-maintenance platform, a number of interactions between platform's components and between users and platform will be performed. As presented above, PM and Coordinator are the main components in the platform core. These components manage the conduct of the maintenance process in accordance with what it was defined in the knowledge base. The TM component saves all traces of these interactions in the knowledge base while respecting the ontological model of process under the domain maintenance ontology.

Hence, after recording these interactions, periodically, depending on the number of instances recorded in the knowledge base the TM analysis this knowledge to extract new knowledge to increase the intelligence level of the platform. Rules resulting from the analysis done by the TM are transmitted to the PM. This later updates the knowledge base in accordance with these rules which allows to the PM to self-manage some special cases of processes. Details about how these two components (TM and PM) are presented in the next section.

To summarize, the first step in these processes is to acquire real knowledge about platform execution, then we apply the learning process, and finally exploit the learning results to adapt the platform's behaviors.

44.4 Motivating Example

44.4.1 Process Model in Maintenance Ontology

In previous work we provided a domain ontology of maintenance to be shared and integrated in the maintenance platform as well as its integrated applications. The ontological model is built in collaboration with maintenance experts; this ontology encloses different models classified as unavoidable in the maintenance scope [14]. The ontology had been conceptualized with UML class diagram and then formalized and implemented by PowerLOOM [15] a logic-based representation language to be exploited in the knowledge base of the s-maintenance platform. Figure 44.3 shows the UML ontological model of the process part of the maintenance ontology. This part presents the main knowledge exploited to setup the self-learning and self-management processes. The description of these concepts is presented in the Appendix A (Fig. 44.2).

44.4.2 Conditional Maintenance Process

As mentioned above, the knowledge base of the s-maintenance platform is composed by the concepts and relations of the ontology as well as their instances. Hence, the presented model in Fig. 44.2 is included in the knowledge base with its instances present the knowledge base of the s-maintenance platform. In our use case, "Conditional Maintenance" is an instance of the concept *Maintenance-type* which is a sub-concept of Maintenance pattern. Also, all steps and transitions of "Conditional Maintenance" are instantiated.

As shown in the presented model, each step can reference an elementary Task or Process Pattern which is a sub-concept of the concept Task. In Fig. 44.3 we present a schematic illustration of the instantiation of the *Process Pattern* concept. We make a zoom on *Management Process* (presenting the process pattern of work request process) which references a step in the *Conditional Maintenance* (instance of Process Pattern). In Appendix B we present the PowerLoom instructions in the knowledge base.

We note that PowerLoom has a static and dynamic query optimizer that is similar to optimizers used in database systems. The dynamic optimizer operates for each conjunctive sub-goal based on actual bindings. Given this mechanism it is

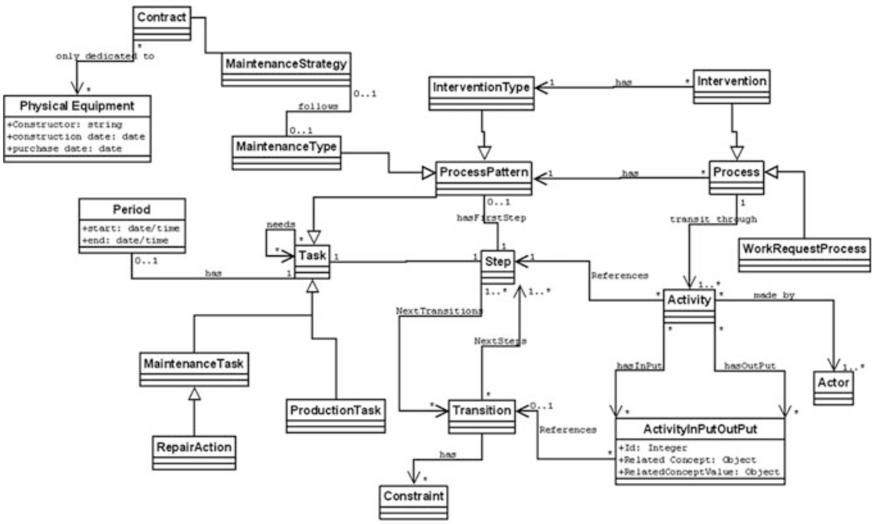


Fig. 44.2 Maintenance ontology: processes model

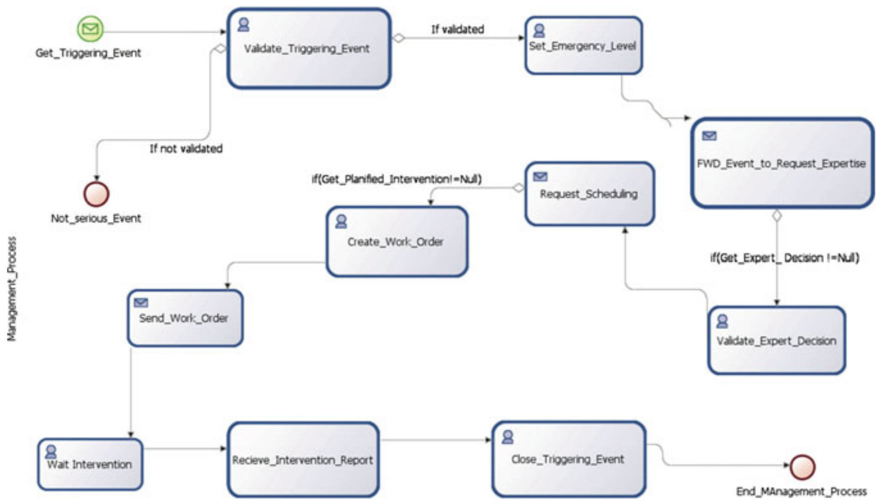


Fig. 44.3 An illustrative workflow of management process (referencing work request process)

```
J48graft pruned tree
------------------

ActivityInPut_Value <= 50
|   Constraint = Alarm.Measure.Value>42
|   |   ActivityInPut_Value <= 43: not_valid (16.0)
|   |   ActivityInPut_Value > 43: valid (4.0)
|   Constraint = Alarm.Measure.Value<42: not_valid (0.0)
|   Constraint = Alarm.Measure.Value=42: not_valid (0.0)
|   Constraint = Alarm.Measure.Value>50: not_valid (14.0)
ActivityInPut_Value > 50: valid (48.0)

Number of Leaves  :     6

Size of the tree :      9
```

Fig. 44.4 An example of learning result rules

possible to run PowerLoom queries that return hundreds of thousands of solutions. PowerLoom also has a powerful relational database interface that allows it to utilize the power of databases for handling large assertion bases [15]. These advantages allow the creation of different views like in databases thanks to some queries jointing concepts and relations. A view consists of a stored query accessible as a virtual table composed of the result set of a query [16]. Table 44.1 presents an example of a view joining each “Maintenance Type” with its “First Step” as well as the “Transition” between the first step and the second step (“Next Step”) in this process pattern.

Table 44.1 Example of maintenance type pattern

Maintenance type	First step	Transition	Next step
Conditional maintenance	Monitoring process	Alarm	Management process
Predictive maintenance	Prognostic	RUL	Management process
Systematic maintenance	Scheduling	Notification (Date)	Management process

In this work we will focus on Management Process in Conditional Maintenance to illustrate the functioning of auto-learning and the auto-management processes in the s-maintenance platform.

44.4.3 Sites and Equipment to be Managed

Our use case concerns Cegelec. This latter is an international firm that designs, installs, and maintains systems in industry. Cegelec has different machines parks distributed in France. Let's consider that Cegelec uses the s-maintenance platform to manage maintenance on these different sites. In these different parks, machines are generally from the same category and the critical supervised components are almost the same. S-maintenance platform takes into account all the equipments and all its components but in this use case we focus only pusher and puller in machine tool. This use case will be treated in the scope of the Process Pattern Conditional Maintenance presented above.

44.5 Functioning of the Self-Learning Process in the Platform

The learning process in the platform will focus this set of concepts *Process Pattern*, *Process, Step*, *Activity*, *Transition* and *Constraint* as well as different relations between these concepts and their instances. So, it will be allow the understanding of different interactions and relations to conclude some dependencies to generate new process patterns. This new knowledge will be then exploited in the self management process to adjust platform's behavior.

To set up this process of self-learning, from different techniques of machine learning [7] we adopt the decision tree approach. According to Nilsson, a decision tree is a tree whose internal nodes are tests (on input patterns) and whose leaf nodes are categories (of patterns). A decision tree assigns a class number (or output) to an input pattern by filtering the pattern down through the tests in the tree. It is to note that there are different tools that transform the decisional tree to decisional rules. Each test has mutually exclusive and exhaustive outcomes. Several systems for learning decision trees have been proposed. Prominent among these are ID3 and its new version, C4.5 [17, 18], CART [19] and incremental methods [20]. From these different systems we adopt the C4.5 algorithm while it seems the most adapted to our needs which can considered as a classification problem.

So, C4.5 is a program that creates a decision tree based on a set of labeled input data. This decision tree can then be tested against unseen labeled test data to quantify how well it generalizes. The C4.5 system consists of four principal programs: (1) decision tree generator ('c4.5') which constructs the decision tree, (2) production rule generator ('c4.5 rules') which is a form production rules from unpruned tree, (3) decision tree interpreter which classifies items using a decision tree, and (4) production rule interpreter which classifies items using a rule set.

The c4.5 outputs are the decision tree proper (weighted training examples/ weighted training error), tables of training error and testing error, and the confusion matrix.

In the case of the proposed architecture of s-maintenance platform, this algorithm is one of the different algorithms included in the "Reasoner" component.

As mentioned by Xu et al. [9] a learning process should satisfy consistency condition that means learning result should be consistent with the original knowledge. This learning process is mainly based on the deep knowledge model and in the view of the knowledge base of the s-maintenance platform. The first step in the learning process is to create data views containing the necessary knowledge to be analyzed from the knowledge base. This step is done in order to present the needed knowledge in a structured form (e.g. table with specific schema). Then, from these views, we create automatically the different files needed by C4.5. The third step is to run the C4.5 algorithm by the "Reasoner" component. Finally, depending to the percentages of the classification correct as well as the relative error percentages, Reasoner, validate or not the learning results.

44.5.1 Views Creation

Like shown in the presented ontological model, each "Process" references a "Process Pattern" and each "Activity" references a "Step". Instances of Process and Activity are the real interactions done in the s-maintenance platform and tracked by the TM component. In addition, each activity is ensured by one or more actors which are defined by instances of Actor concept.

In reality, the knowledge that will be the base of the learning exercise is the instances of Activity and/or Process concepts. Although, in order to have a generalized rules for the learning results, we don't applied the learning algorithm on these instances because all of these instances are different but we up a notch. So, we apply the learning on concepts level. In a first time we built a view containing instances, then in a second time we generate a second view containing instances of general concepts, each instances of Activity or Process is replaced by the instance that it references on the concepts of Process Pattern or Step. Tables 44.2 and 44.3

Table 44.2 First level view of learning process activity

Input id	Activity	Activity output	Next activity id	Related physical equipment
234	A982P22	536	A1236P22	P456
536	A1236P22	589	A2356P22	P456
789	A254P11	465	A269P11	Pu342
865	A11P18	965	A36P18	I231
965	A36P18	136	A39P18	I231
136	A39P18	245	A85P18	I231

presents an explicative example about this views creation. It is to note that, views schemas are previously designed in the system respecting the ontological model presented above. The proposed schemas are inspired from the state transition table which is a table showing what state a finite state machine will move to, based on the current state and other inputs [22]. The first view's schema is presented by the 5-uplet: <ActivityInPut.Id, Activity, ActivityOutPut.Id, NextActivity, Related-PhysicalEquipment> as shown in Table 44.2.

The second schema is mainly composed by the following 8-uplet *<ActivityInPut.Concept, ActivityInPut.Val, Step, ActivityOutPut.Concept, ActivityOutPut.Val, Constraint, NextStep, ConcernedEquipmentModel>*. This second view is generated based on the first one. *Step* and *NextStep* in the second view presents the referenced *Step* of *Activity* and *Next Activity* in the first view. *ActivityInPut.Concept* and *ActivityInPut.Val* presents values of the other attributes of *ActivityInPut* instances and the same thing for *ActivityOutPut* instances. Concerning *Constraint*, it is generated from the *Transition* referenced by the *ActivityInPutOutPut* concept. *ConcernedEquipmentModel* presents "*Equipment Model*" concept which is an abstract model of *"Physical Equipment"* concept. Each Physical Equipment has an Equipment Model.

44.5.2 C4.5 Files Creation

C4.5 expects to find 3 files: "filestem.names", "filestem.data", "filestem.test". The input to C4.5 consists of a collection of training cases, each having a group of values for a fixed set of attributes (or independent variables) A = {A1, A2, … An} and a class attribute (or dependent variable). An attribute Aa is described as continuous or discrete according to whether its values are numeric or nominal. The goal is to learn from the training cases a function that maps from the attribute values to a predicted class [23].

Hence, our data sets of learning must be well structured by choosing the best set independent variables and the dependent variable. Then, C4.5 files "filestem.names", "filestem.data" must be well edited to be used in the learning process. The .names file describes the dataset, while the .data and .test files contain the examples which make up the dataset. Like mentioned above, our learning process will be focused on activities results. When speaking about activities we

Table 44.3 Second level view of learning process

Activity input concept	Activity input val	Step	Activity output concept	Activity output val	Constraint	Next step	Concerned equipment model
Alarm measure value	46	Validate alarm	Alarm validated	True	Alarm measure value > 42	Set emergency level	Pusher
Alarm validated	True	Set emergency level	Alarm emergency	Very high	Alarm. measure value > 42 and alarm validate = True	Request expertise	Pusher
Alarm	44	Validate alarm	Alarm validated	False	Alarm value > 42	Close work request	Puller
Measure	41	Control data	Condition	Not violated	Measure value > 42	Control data	Indexer
Measure	42	Control data	Condition	Not violated	Measure value > 42	Control data	Indexer
Measure	43	Control data	Condition	Violated	Measure value > 42	Create alarm	Indexer

can move to speak about steps as done in the views creation. In order to well structure C4.5 files, we create a couple of files (".names", ".data") for each instance of Step concept. The attributes of these files are chosen from the second view. Therefore, while the learning will be done on activities output, in all files the *ActivityOutPut.Value* will be the dependent variable and the set of independent variable will contains {*ActivityInPut.Concept, ActivityInPut.Value, ActivityOutPut.Concept, Constraint, ConcernedEquipmentModel*}.

44.5.3 C4.5 Running

As shown in the auto-learning and auto-management processes architecture presented in Fig. 44.4, the TM component requests the Reasoner component to launch the learning algorithm.

Regarding the launching of C4.5 algorithm, two choices was presented; either periodically (e.g., every 1 month), or launched it according to threshold on number of instances in the knowledge base (e.g., the algorithm will be launched every 100 new instances).

From these two propositions, we chose the latter one because the learning result depends mainly on the number of instances contained in the dataset to be analyzed. Generally, the number of instances does not vary consistently or according to unified period. In other terms, we can have for example only 9 instances during one month and 100 instances in two days. So it is better to launch of the algorithm depending to the number of instances. This later will be checked by the TM every time a new instance is added to the knowledge base.

44.5.4 Results Validation

To validate the extracted rules and taking into account the learning result or not by the platform, we have a defined a trust level according to the classification criteria.

Correctly Classified Instances, Kappa statistic, Relative absolute error, and relative squared error are the main indicators that we used to validate the learning results. The correctly and incorrectly classified instances show the percentage of test instances that were correctly and instances classified. Kappa is a chance-corrected measure of agreement between the classifications and the true classes. It's calculated by taking the agreement expected by chance away from the observed agreement and dividing by the maximum possible agreement. The error rates are used for numeric prediction rather than classification. In numeric prediction, predictions aren't just right or wrong, the error has a magnitude, and these measures reflect that [21].

Table 44.4 Validation indicators' thresholds

Indicator	Threshold
Correctly classified instances	>=90 %
Kappa statistic	>=0.5
Relative absolute error	<=10 %
Relative squared error	<50 %

While management and decision-making in maintenance processes are very critical tasks, especially if these decisions will be done automatically and will enables the platform of maintenance to self-manage processes, we adopted a strict validation policy by defining very high trust thresholds as shown in Table 44.4. Also, we note that all of the thresholds must be respected, to validate the learning results.

44.6 Function of the Self-Management Process in the Platform

The main functionality of the PM is to update the Knowledge base (KB) by the learning results and ensure the automatic execution of some activities in the workflow of the maintenance process without human intervention. But, before updating the KB, learning result must be adapted to the KB structure, hence respect the ontological model of process. Some transformations must be done by the PM. It is to note that this update can take different ways: (1) adding new rules and keeping original rules; (2) adding new rules and modifying some original rules; and (3) only modifying original rules. In addition, this update will mainly concern the instances of *step, transition* and *constraint* concepts. A modification in this level of the ontology causes a change in the workflow of process patterns. Hence, this process is composed by three steps, firstly adapt learning rules, secondly update the KB and finally self-manage the new added processes.

44.6.1 Learning Rules Adapting

Figure 44.5 shows an example of a learning result rules. The PM must translate these rules to the process ontology view form. The learning result is composed by rules defined by the attributes defined in the dataset. The PM takes advantages from the fact that the structure of all dataset files is the same for all instances of Step concept and that the attributes presents the ontology concepts. So the resulting rules are expressed using the ontology concepts and their instances or related terms.

```
(DEFRELATION Confirmed-Self-Managed((?C Constraint) (?Self-Managed Boolean)))
(Assert (Step Validate-Alarm))
(Assert (Transition Alarm-Not-Valid))
(Assert (Transition Alarm-Valid))
(Assert (Constraint Alarm.Measure.Value>42))
(Assert (Constraint Alarm.Measure.Value<42))
(Assert (Constraint Alarm.Measure.Value=42))
(Assert (has-constraint Alarm-Not-Valid  Alarm.Measure.Value<42))
(Assert (has-constraint Alarm-Valid  Alarm.Measure.Value>42))
(Assert (has-constraint Alarm-Not-Valid  Alarm.Measure.Value=42))
(Assert (Confirmed -Self-Managed  Alarm.Measure.Value=42 False))
(Assert (Confirmed -Self-Managed  Alarm.Measure.Value<42 False))
(Assert (Confirmed -Self-Managed  Alarm.Measure.Value>42 False))
```

Fig. 44.5 Initial KB

Table 44.5 Interpretation table

Step	Extracted constraint	Learned transition	Existed constraint
Validate-alarm	Alarm measure value > 50	Alarm-valid	Alarm measure value > 50
Validate-alarm	Alarm measure value <= 43	Alarm-not-valid	Alarm measure value > 42
Validate-alarm	Alarm measure value > 43	Alarm-valid	Alarm measure value > 42
Validate-alarm	Null	Null	Alarm measure value < 42
Validate-alarm	Null	Null	Alarm measure value = 42
Validate-alarm	Alarm measure value <= 50	Alarm-not-valid	Alarm measure value > 50

In the following example provides rules about possible transitions and their relative constraints related to "Validate Alarm" an instance of Step concept.

The result obtained need to be analyzed by using the second level view related to the analyzed Step. This level view permits to understand the relation between attributes used in the rules and concepts. For example, regarding to the second level view, the ActivityInPutValue in the case of "Validate Alarm" step presents the value of Alarm.Measure.Value. To interpret the learning result, the PM construct an interpretation table having the as schema the 4-uplet (Step, Extracted Constraint, Learned Transition, Existed constraint). Table 44.5 presents the interpretation table of the learning result about "Validate alarm" step.

From this table, the PM interprets rules:

- *Ignore* contradictory rules
  - The extracted constraint Alarm.Measure.Value <= 50 is contradictious with the existed constraint Alarm.Measure.Value > 50.
- *Modify existing rules*: Alarm.Measure.Value > 42, Alarm.Measure.Value < 42, Alarm.Measure.Value =42 will be modified by Alarm.Measure.Value <= 43 and Alarm.Measure.Value > 43.

From the interpreted rules, *tow action must be done*

- Remove relations with the existing constraints to be modified
- Add new relations for the learned rules

- Rule 1: If Alarm.Measure.Value > 43 then automatically Transition Valid_Alarm
- Rule 2: If Alarm.Measure.Value <= 43 then automatically Transition Not_Valid_Alarm
- Rule 3: If Alarm.Measure.Value > 50 then automatically Transition Valid_Alarm

After that, rules verification, this step is to analyze if there is contradictious rules or if there is a rule is more general than another one. For example, Rule 1 is more generic than Rule 3, so the more general rule will be adopted. Consequently, in the end of this step two rules are validated Rule 1 and Rule 2 with rules to be modified.

44.6.2 KB Updating

The three concepts *step, transition* and *constraint* are considered as the representative concepts of the abstract models of the maintenance processes' workflows. The other parallel concepts Process, Activity and ActivityInPutOutPut are the concepts representing the real and physical workflows instantiated during the execution of the platform. For this reason, the update of the knowledge base concerns instances of step, transition and constraint concepts. To explain how the PM updates the KB we show the initial KB and the resulting knowledge base. Associations and attributes of classes are presented by "`DEFFUNCTION`" and "`DEFRELATION`" commands. The command "`ASSERT`" is used to define different instances of concepts and relations.

Attributes of the concepts are presented as relations between the concept and the attribute name and its type. As shown in Fig. 44.6 of the initial KB, `Confirmed-Self-Managed` is a Boolean attribute of the constraint concept. This attribute indicates if the constraint is validated to be self managed or not. So one of the first thing to be done by the PM when updating the KB is a constraint is validated to be self-managed is to change the value of this attribute. Also, according the learned rules, the PM begins by retracting instances to be deleted, and then add the new instances. The KB updated in Fig. 44.7 shows the retracted relations between the transitions Alarm-Valid and Alarm-Not-Valid with some constraints, the instantiation of the new constraints and their relations with the

```
Modify Existing Rules
(Retract (has-constraint Alarm-Valid  Alarm.Measure.Value>42))
(Retract (has-constraint Alarm-Not-Valid  Alarm.Measure.Value<42))
(Retract (has-constraint Alarm-Not-Valid  Alarm.Measure.Value=42))
(Assert (Constraint Alarm.Measure.Value>43))
(Assert (Constraint Alarm.Measure.Value<=43))

Add New Rules
(Assert (has-constraint Alarm-Valid  Alarm.Measure.Value>43))
(Assert (has-constraint Alarm-Not-Valid  Alarm.Measure.Value<=43))
(Assert (Confirmed-Self-Managed  Alarm.Measure.Value<=43 True))
(Assert (Confirmed-Self-Managed  Alarm.Measure.Value>43 True))
```

Fig. 44.6 KB updated by the PM

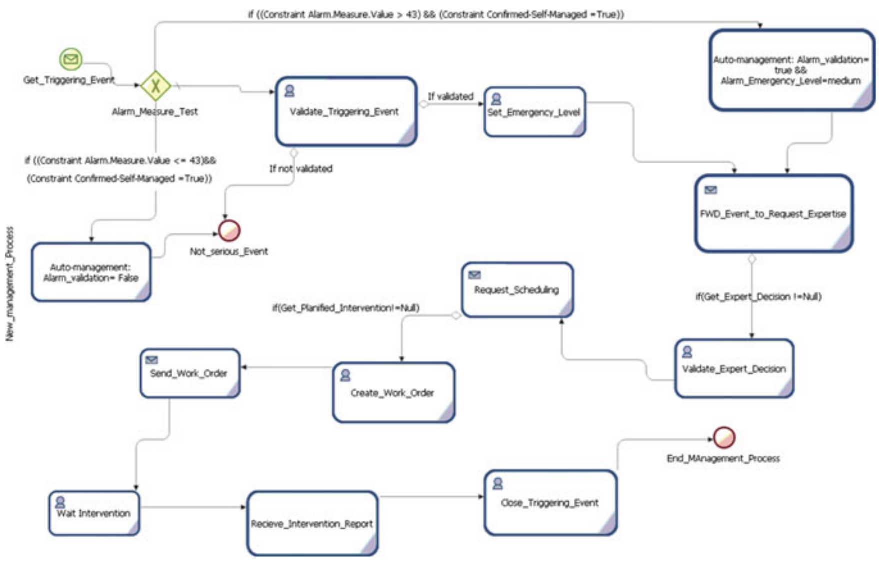


Fig. 44.7 New version of management process's workflow

transitions Alarm-Valid and Alarm-Not-Valid and finally the setting of the value true to the `Confirmed-Self-Managed` attributes of these constraints.

After updating the knowledge base, the workflow of the process pattern is updated too. Figure 44.8 shows the new version of the workflow of management process different from the first workflow shown in Fig. 44.4.

44.6.3 Self-Management of New Processes Execution

When running the platform, during the progress of a particular process (an instance of the concept process) that follows a pattern process, activities related to the various steps of this process pattern are instantiated. These activities are created and performed by actors (human or software).

After the self-learning process, activities (instances of Acitivity concept) as well as the inputs and outputs of these activities (instances of AcitivityInPutOutPut concept) will be self-managed (e.g., instantiated automatically) by the PM. Figures 44.9 and 44.10 shows the difference between the execution before self learning when execution are ensured by actors and execution after self-learning with self-management ensure by the PM. Information presented in these figures are extracted from the Knowledge base of the s-maintenance platform. In Fig. 44.9 activities are created when actors interact with the platform and the transition from an activity to the next one is also done by actors. Contrary, in Fig. 44.10, activities and transitions are created automatically by the PM while it previously knows the results of the activities thanks to the learned rules (i.e. the new knowledge base).

```
Initial KB:

(Assert (Step Validate-Alarm))
(Assert (Transition Alarm-Not-Valid))
(Assert (Transition Alarm-Valid))
(Assert (Transition Alarm.Measure))
(Assert (Constraint Alarm.Measure.Value>42))
(Assert (Constraint Alarm.Measure.Value<42))
(Assert (Constraint Alarm.Measure.Value=42))
(Assert (has-constraint Alarm-Not-Valid  Alarm.Measure.Value<42))
(Assert (has-constraint Alarm-Valid  Alarm.Measure.Value>42))
(Assert (has-constraint Alarm-Not-Valid  Alarm.Measure.Value=42))
(Assert (Confirmed -Self-Managed  Alarm.Measure.Value=42 False))
(Assert (Confirmed -Self-Managed  Alarm.Measure.Value<42 False))
(Assert (Confirmed -Self-Managed  Alarm.Measure.Value>42 False))

Platform Actors' interactions:

(Assert (Activity A235))
(Assert (has-Reference-step A235 Validate-Alarm))
(Assert (ActivityInPutOutPut AIO569))
(Assert (has-Reference-Transition AIO569 Alarm.Measure))
(Assert (has-Input A235 AIO569))
(Assert (ActivityInPutOutPut AIO589))
(Assert (has-Reference-Transition AIO589 Alarm-Valid))
(Assert (has-Output A235 AIO589))
```

Fig. 44.8 Execution before self-learning and without self-management

```
New KB:

(Assert (Step Validate-Alarm))
(Assert (Transition Alarm-Not-Valid))
(Assert (Transition Alarm-Valid))
(Assert (Transition Alarm.Measure))
(Assert (Constraint Alarm.Measure.Value>43))
(Assert (Constraint Alarm.Measure.Value<=43))
(Assert (has-constraint Alarm-Valid  Alarm.Measure.Value>43))
(Assert (has-constraint Alarm-Not-Valid  Alarm.Measure.Value<=43))
(Assert (Confirmed -Self-Managed  Alarm.Measure.Value<=43 True))
(Assert (Confirmed -Self-Managed  Alarm.Measure.Value>43 True))

PM: self management: automatic assertions

(Assert (Activity A235))
(Assert (has-Reference-step A235 Validate-Alarm))
(Assert (ActivityInPutOutPut AIO569))
(Assert (has-Reference-Transition AIO569 Alarm.Measure))
(Assert (has-Input A235 AIO569))
(Assert (ActivityInPutOutPut AIO589))
(Assert (has-Reference-Transition AIO589 Alarm-Valid))
(Assert (has-Output A235 AIO589))
```

Fig. 44.9 Execution after self-learning and with self-management

44.7 Conclusion

E-maintenance and s-maintenance are considered as the most important resulting concepts that have been developed and applied in the industry. The difference between these two concepts is in the level of provided services. Services provided by s-maintenance are considered more dynamic than those provided by e-maintenance. S-maintenance is mainly characterized by its self-learning and self-management capacities. In previous work we proposed an architecture of an s-maintenance platform permitting to ensure these characteristics.

In this paper we focus the functioning of the self-learning and self-management processes in the proposed platform. The two processes are mainly based on the maintenance domain ontology that we proposed in previous work and especially the part of ontology conceptualizing maintenance processes. The maintenance ontology is presented on PowerLOOM an expressive description language in the knowledge base KB of the s-maintenance platform. The self-learning process uses the machine learning algorithm C4.5. This algorithm is applied on the knowledge saved from platform execution traces ensured by the TM component. After that the learning results will be exploited by the PM component which ensures the self-management process by the update of the KB according to the learned rules and self-manages some activities in the platform execution that were ensured by the platform actors (users or integrated applications). The explanation of the functioning of these processes is accompanied with illustrative examples.

On the other hand, we note that the success of the s-maintenance platform and especially the self-learning and self-management processes depends to the number of the managed equipment while these processes are based on the tracked knowledge about the interactions of users with platform and the executions of the maintenance processes. Hence, it depends also to time, we cannot predict on how much time the learning process can be launched and if the result will be validated or not. For this reasons, the platform must be tested in large scale. Another problem can be confronted, is identification of the threshold of knowledge to launch the process of learning. This point must be treated with experts to identify these thresholds to obtain good learning results. In addition, another question must be treated in future work is the question about how we can evaluate the impact factor of these processes on the platform functioning as well as and especially on maintenance in the industrial scope. In other terms, how we can quantify the gain in terms of time and cost in the company.

Acknowledgments This work was carried out and funded in the framework of SMAC project (Semantic-maintenance and life cycle), supported by Interreg IV program between France and Switzerland.

Appendix A

See Table 44.6

Table 44.6 Data dictionary of the processes model *concept name*

	Description
Physical equipment	Tangible, instantiated, serialized object, component, device, subsystem, functional unit, equipment or system that can be individually considered to be maintained. A physical equipment may be an entire facility, an entire functioning platform (such as an CH-47 Tail Number XYZ helicopter), or a component piece of equipment, such as a specific instance of a bearing
Maintenance strategy	A long-term plan, covering all aspects of maintenance management which sets the direction for maintenance management, and contains firm action plans for achieving a desired future state for the maintenance function respecting a maintenance type
Process	Sequence of interdependent and linked activities which, at every step, consume one or more resources (employee time, energy, machines, money) to convert inputs (data, material, parts, etc.) into outputs while respecting a process pattern (i.e. the process pattern presents the general model of the process). These Transition outputs then serve as transition inputs for the next step until a known goal is reached
Intervention	Specific type of process concerning the main technical part of maintenance which is the core of maintenance
Process pattern	Pattern which describes a proven, successful approach and/or series of actions. A pattern is a description of a general solution to a common problem or issue from which a detailed solution to a specific problem may be determined
Maintenance type	Specific type of process pattern concerning the method to carry out maintenance and the orchestration of processes used in order to achieve the maintenance strategy goals
	There are 12 possible types of maintenance which are: preventive maintenance, scheduled maintenance, predetermined maintenance, condition based maintenance, predictive maintenance, corrective maintenance, remote maintenance, deferred maintenance, immediate maintenance, on line maintenance, on-site maintenance and operator maintenance
Intervention type	Specific type of process pattern. It is the principle method of conveying the appropriate activities to all parties involved in an intervention on a physical equipment. It presents the generic model of an intervention
Task	Piece of work assigned or done as part of one's duties. The task is evaluated by looking at its outcome in terms of completeness, accuracy, tolerance, clarity, error, or quantity
Transition	Passage from one step to another in a process pattern while the actual step is finished. It is considered as the event that launches the next step. Also it is considered as output of the actual step or and the input of the next step

(continued)

Table 44.6 (continued)

	Description
Constraint	Restrictive condition to control transitions between steps
Activity	Organizational unit to perform a specific action ensured by an actor
Activity input output	Real transition between activities in the system. Generally it references Transitions and presents the input and output of activities but not in all cases like in the case of the first step in a process pattern
Step	Maneuver made as part of progress toward the progress of a process, referenced by an activity
Work request process	Specific type of process launched automatically or by an actor when receiving a triggering event. It allows managing the work request until resolving the origin problem of the triggering event and the end of the intervention process including the intervention report's edition
Contract	Specific type of document defining a service agreement between a service provider and a physical equipments owner or exploiter to maintain a set of physical equipment

Appendix B

The following PowerLOOM instructions present the Knowledge base containing the part of the ontology shown above concerning process management. Each UML class is translated into a PowerLOOM concept using the "DEFCONCEPT" command. Associations and attributes of classes are translated into a PowerLOOM relation or function using the "DEFFUNCTION" and "DEFRELATION" commands. The command "ASSERT" is used to define different instances of concepts and relations.

```
(DEFCONCEPT Task)
(DEFCONCEPT Process-Pattern (?T Task))
(DEFCONCEPT Maintenance-Type (?PT Process-Pattern))
(DEFCONCEPT Step)
(DEFCONCEPT First-Step (?S Step))
(DEFCONCEPT Transition)
(DEFCONCEPT Next-Step (?S Step))
(DEFCONCEPT Constraint)
(Defrelation (StepInPut ((?T Transition) (?NS Step)))
(Defrelation (hasFirstStep ((?MT Maintenance-Type) (?FS
First-Step)))
(Defrelation (StepOutPut ((?FS Step) (?T Transition)))
(Defrelation (Refrences ((?S Step) (?T Task)))
(Assert (Maintenace-Type Conditional-Maintenance)
(Assert (Maintenace-Type Predective-Maintenance)
(Assert (Maintenace-Type Systematic-Maintenance)
(Assert (Maintenace-Type Systematic-Maintenance)
(Assert (Step Monitoring-Process)
```

```
(Assert (Step Prognostic)
(Assert (Step Scheduling-Process)
(Assert (Management-Process)
(Assert (Transition Alarm)
(Assert (Transition RUL)
(Assert (Transition Notification-Date)
(ASSERT (hasFirstStep Conditional-Maintenance Monitor-
ing-Process))
(ASSERT (hasFirstStep Predective-Maintenance
Prognostic))
(ASSERT (hasFirstStep Systematic-Maintenance Scheduling-
Process))
(ASSERT (StepOutPut Monitoring-Process Alarm))
(ASSERT (StepOutPut Prognostic RUL))
(ASSERT (StepOutPut Scheduling-Process Notification-
Date))
(ASSERT (StepInPut RUL Management-Process))
(ASSERT (StepInPut Alarm Management-Process))
(ASSERT (StepInPut Notification-Date Management-Process))
```

References

1. Li Y, Chun L, Nee A, Ching Y (2005) An agent-based platform for web enabled equipment predictive maintenance. Proceedings of IAT'05 IEEE/WIC/ACM international conference on intelligent agent technology, Compie'gne, France
2. Chebel-Morello B, Karray M H, Zerhouni N (2010) New perspectives of Maintenance systems: "towards s-maintenance". In: Proceedings of the IMS2010 Summer School on sustainable manufacturing. Zurich
3. Karray MH, Chebel-Morello B, Lang C, Zerhouni N (2011) A component based system for S-maintenance. In: Proceedings of 9th IEEE International Conference on industrial informatics, pp 519–526
4. Ganek AG, Corbi TA (2003) The dawning of the autonomic computing era. IBM Syst J 42(1):5–18
5. Leondes CT (2000) Knowledge—based systems: techniques and applications. Academic Press
6. Poole D, Mackworth A, Goebel R (1998) Computational intelligence: a logical approach. Oxford University Press
7. Nilsson J (1998) Introduction to machine learning: an early draft of a proposed textbook. Available from http://robotics.stanford.edu/people/nilsson/mlbook.html, pp 175–188
8. Robin (2010) Machine learning overview. Available from http://intelligence.worldofcomputing.net/machine-learning/machine-learning-overview.html
9. Xu CS, Xu ZM, Xiao PD, Zhou ZY, Liu SX, Jiang ZH (1995) A self-learning system and its application in fault diagnosis. In: Proceedings of the instrumentation and measurement technology conference
10. Lowe G, Shirinzadeh B (2005) A knowledge-base self-learning system. In: Proceedings of the (483) ACIT–automation, control, and applications

11. Ganek AG, Corbi TA (2003) The dawning of the autonomic computing era. IBM Syst J 42(1):5–18
12. Haydarlou AR, Oey MA, Overeinder BJ, Brazier FMT (2006) Using semantic web technology for self-management of distributed object-oriented systems. In: Proceedings of the 2006 IEEE/WIC/ACM international conference on web intelligence (WI-06). Hong Kong, China
13. http://en.wikipedia.org/wiki/Workers%27_self-management 2010
14. Karray MH, Chebel-Morello B, Zerhouni N (2011) A formal ontology for industrial maintenance. In: Proceedings of the 5th terminology & ontology: theories and applications conference 2011
15. Chalupsky H, MacGregor RM, Russ T (2010) PowerLOOM manual: powerful knowledge representation and reasoning with delivery in Common-Lisp, Java, and C++ Version: 1.48
16. http://en.wikipedia.org/wiki/View_%28database%29
17. Quinlan JR (1986) Induction of decision trees. Mach Learn 1(1):81–106
18. Quinlan JR (1993) C4.5: programs for machine learning. Morgan Kaufmann, San Francisco
19. Breiman L, Friedman J, Olshen R, Stone C (1984) Classification and regression trees. Chapman and Hall/CRC, Monterey, CA
20. Utgoff P (1989) Incremental induction of decision trees. Mach Learn 4(2):161–186
21. Bouckaert RR, Frank W, Hall M, Kirkby R, Reutemann P, Seewald A, Scuse D. WEKA Manual for Version 3-6-4. The University of Waikato
22. Ml Breen (2005) Experience of using a lightweight formal specification method for a commercial embedded system product line. Requirements Eng J 10(2):161–172. doi:10.1007/s00766-004-0209-1
23. Kohavi R, Quinlan (2002) Decision-tree discovery. In: Klosgen W, Zytkow JM (eds) Handbook of data mining and knowledge discovery. Oxford University Press, Chapter 16.1.3, pp 267–276

Chapter 45
Case Study: Level of Service Criteria for Critical Rotating Assets

S. Haines and M. R. Hodkiewicz

Abstract This case study describes how the process of developing an asset management plan, adapted from the infrastructure industry, is used to translate organizational objectives to tangible operational, reliability, contractual, and regulatory objectives at the asset level. In this case, the assets are gas-turbine or reciprocating compressors critical to the organizational performance of a gas transmission company. The key to this process is the establishment of 'level of service' (LOS) criteria against which asset performance and the performance of the processes that support their operation are assessed. These LOS criteria, their measures and targets, are explicitly linked to the organization's business plan and aligned with key policies. There are specific challenges to selecting performance measures associated with compressors in gas transmission lines and these are discussed. The process allows for performance comparison of the 68 compressors, supporting risk identification, operational improvements, and knowledge sharing. This view of a single asset class across the business units 'horizontal asset management' also identifies common risks and opportunities for technical improvement. The business case for investment in technical projects such as prognostics is enhanced as success can be leveraged across a number of critical units in the business and performance improvement assessed against common measures.

S. Haines (✉) · M. R. Hodkiewicz
University of Western Australia, Perth, Australia
e-mail: samantha.haines@uwa.edu.au

M. R. Hodkiewicz
e-mail: melinda.howkiewicz@uwa.edu.au

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_45,

45.1 Introduction

Despite the advances made in organizations to align operating units activities with organizational strategic objectives through tools such as the balanced score card, too often units fail to coordinate [1]. As strategic objectives are translated down through the levels, from corporate, to senior management, to team management and finally to the team members, who are at the forefront of day-to-day decision making, maintaining alignment is a challenge. This results in "performance-sapping disagreement, lost opportunities, wasted resources [2]."

In the last decade Asset Management (AM) has emerged as its own knowledge field and is increasingly visible in organizational structures and corporate decision-making. Regulatory bodies across the world are requiring Pipeline Owners and Operators to establish formal AM systems. The most widely used requirements-based framework is PAS 55 [3], a Publically Available Specification for management of physical assets. The UK regulator for electricity and gas operators, Ofgem, actively encourages certification with PAS 55, and is now auditing companies with an approach in alignment with PAS 55 [4]. The UK Office of Rail has begun to formally evaluate the national rail network's AM processes with respect to PAS 55 [5] and in Australia, Energy Safe Victoria (ESV), "encourages and supports companies gaining AM accreditation such as to PAS 55" [5]. In March 2011, work started on a new set of ISO Standards (ISO 55000-55002) for Asset Management.

One of the core concepts in PAS 55 is that of "line of sight" between AM activities and organizational objectives. How this is achieved is up to the organization. While the use of the balance score card approach [6] works well to align organizational, business unit, and AM management system strategic objectives, it does not translate well to the physical asset level.

The infrastructure sector has for a long time used a level of service approach to identify measures that align asset-related activities to the achievement of stakeholder levels of service. Horizontal AM planning is defined in this paper as asset management of a class of assets across multiple asset systems. The concept is well established in the municipality and energy industries [7–9] however, documented process on how to achieve a view of common assets across different operating units is not widely available. There is little published work of the use of this approach in the gas pipeline industry and none located showing a suite of expected performance measures aligned with organizational objectives for gas compressors. This paper describes a process to do this in the context of a large national gas transmission organization.

45.2 Levels of Service

A level of service (LOS) is "the defined service quality for a particular activity or service area against which service performance can be measured" [8]. Service levels usually relate to quantity, quality, reliability, environmental acceptability, and cost. The selection of a set of LOS measures and the setting of measures and targets has been integral to infrastructure asset management for decades, starting with the road sector in 1985 [10, 11]. The identification of relevant LOS is the first step in the process of developing AM Plans according to the International Infrastructure Management Manual. Following steps include predicting demand, assess cost, condition and performance, failure mode and risk analysis, evaluate/select treatment options, select the optimum solution, assess financial cash flows, prepare the AM Plan, implement and improve. LOS is widely used in industries as diverse as highways, water and storm water, airport passenger terminals, and local government assets [11–16]. Principles when selecting LOS are that they are (1) linked with organizational strategic drivers, (2) developed in consultation with customers/users and address issues they believe are important, and (3) measurable. Typical organizational strategic drivers include meeting license to operate requirements, safety and environmental targets, maintaining customer satisfaction, achieving production, quality and cost targets, and retaining core people.

The existence of an agreed set of LOS and associated measures provides a target for asset managers. This encourages the asset manager to examine different strategies to deliver the target allowing a focus on a wider set of measures (than the traditional tons and cost) for the operations and maintenance groups to focus on. It encourages operations and maintenance to have a discussion about what they are prepared to trade off to achieve specific goals rather than assume that availability, cost and production can all be optimized.

The AM Plan consolidates in one document the full asset life cycle processes and practises used to ensure optimal asset outcomes. It demonstrates to key stakeholders that the asset management approach is prudent, delivers long term sustainability, addresses an appropriate balance between service levels, performance, cost and risk, provides the technical basis to support expenditure, and provides the basis for continuous improvement of asset management lifecycle management.

45.3 Method

The development of the LOS table is an iterative process with the main stages shown in Fig. 45.1.

The 1st stage is the identification of organizational objectives, stakeholder and business context. These inform the selection of appropriate LOS, measures and targets. Typically business drivers in industries like gas and mining are in the following categories: Safety and Environment; Legislative and Regulatory; Production and/or contractual obligations; and Cost. These need to be translated in consultation with stakeholders into LOS criteria.

Development of LOS for the compressors involved a review of:

1. pipeline safety cases and safety management systems;
2. commercial contracts, and terms and conditions;

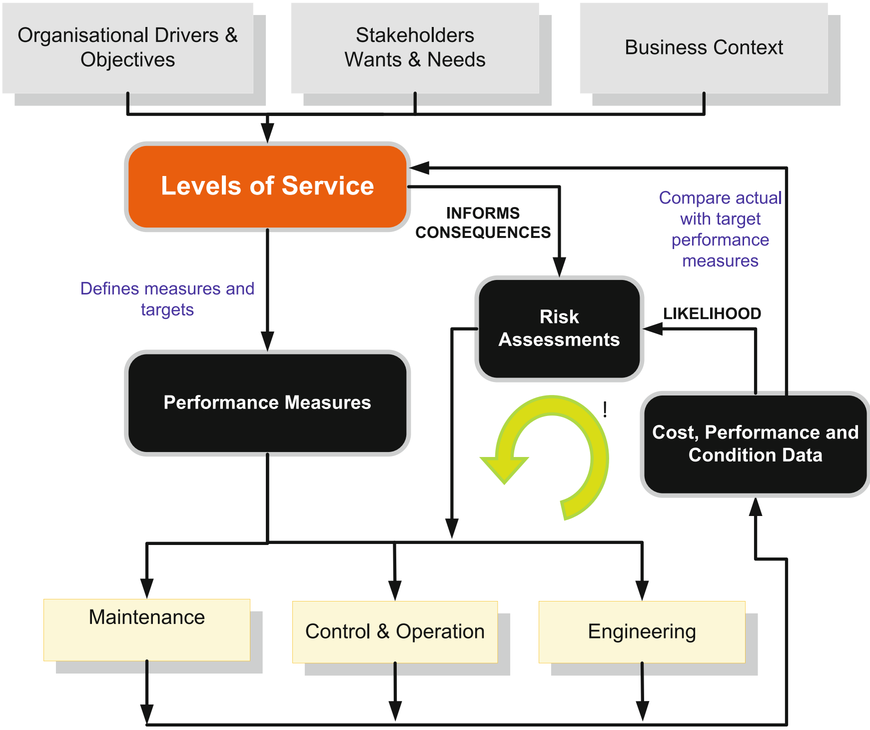


Fig. 45.1 Steps in the development of levels of service

3. asset reports (performance, integrity, technical queries, and technical change);
4. current legislation (Australian Standards [17–19] and State regulations);
5. historical failures and the consequences of failure [20];
6. current and historical operation and maintenance regimes; and
7. current tools and systems used in monitoring the compressors [21–23].

While information on items (1) to (4) is available centrally within the Company, data on the assets (5) to (7) in the different business units is less easy to obtain as each has their own processes and systems. Data was collected via phone or face-to-face interviews and file exchange with Company employees and other industry professionals.

The 2nd stage is deciding what measures are appropriate to monitor progress towards and achievement of the LOS criteria. In some cases a suite of measures is appropriate. Good practice is that across all the LOS criteria there should be no more than 10 measures. Selecting performance measures needs considerable stakeholder input. Sanity checks at this stage include questions such as [24] Truth—is it measuring what it is meant to? Focus—is it measuring only what it is meant to? Consistency—it the measure consistent whenever and whoever measures? Access—can the data be readily communicated and understood? Cost—

how expensive is it to collect, collate, and analyze the data? Gaming—will the measure encourage undesirable behaviors?

The 3rd stage is the risk assessment. This process follows ISO 31000 starts with an agreement about the context(s) and function of the asset, in this case, a compressor and then risk identification and assessment. It is common practice to use the FMEA process for this. This focuses on the events which could result in failure to deliver function and hence the target LOS measures stated in the table.

The 4th stage is risk evaluation and ranking. Use of the Risk Priority Number approach (RPN) consider three dimensions, the likelihood of failure, the consequence of failure, and the controls currently in place.

- The likelihood of functional failure evaluation needs to consider the condition and age of the asset, and potential events that could lead to loss of function.
- The consequence evaluation is linked directly to the LOS; this constrains consideration of the consequence to an inability to meet the targets in the LOS table. Organizations have their own consequence severity tables; these are used to grade the severity of the risk.
- Controls include current maintenance and operating practises, process control, condition monitoring, inspections that can detect incipient failures, and/or prevent escalation of failure events.

The 5th stage involves developing the LOS table which captures the results of the assessment described above.

The 6th and final stage is communicating this with the stakeholders and establishing a process to routinely review the performance of the asset against the desired levels of service targets and the asset level of service targets against the organizational objectives. This review is usually done as part of the annual review of the asset AM Plan and results in a series of actions to close any gaps that might exist between target and actual LOS. The LOS table is the key input to an AM Plan. Placed at the forefront of the AM Plan, it tells the reader succinctly the asset requirements, the reason for the requirements, how and what to measure, the desired target(s), and if the asset/asset class/asset system is performing to target, or not.

45.4 Results

45.4.1 The Current Situation

Natural gas is of growing importance to the energy infrastructure across the globe. In Australia natural gas is a vital part of the energy equation both for domestic use and export; Australia consumed 26.6 billion cu m of natural gas in 2009. The case study company operates 20 natural gas pipelines in 7 states across Australia and has more than $8 billion of gas transmission and distribution assets. There are a number of operating units based on the geographical location of the individual pipelines. The pipelines of interest in this case are the 10,000 km of high pressure transmission lines that move the gas, using compressors, to deliver contracted gas volume and pressure. The compressors are located at compressor stations along the pipeline to maintain supply pressures in the range 1,050–10,000 kPa.

The case study company was amalgamated from a number of smaller companies each with their own operating and maintenance approaches. In 2008/2009 there was considerable focus on consolidating the operations teams to have a national focus aimed at optimizing, streamlining, and standardizing activities to think nationally and act locally. This process included the formation of a national AM team responsible for AM process and procedures, asset risk management, and asset performance reporting. Although the company already produced an AM Plan, there was no explicit focus on compressors. The different sections of the AM Plan are: Transmission Pipelines, Distribution Mains and Services, Gate Stations, Pressure Regulating and Valve Installations, Consumer Metering Installations, Odorant Facilities, and Network Monitoring and Control (SCADA) Facilities. The one mention of compressors is a line item identifying the need to do preventative maintenance on the compressors to 'reduce the probability of failure', there was therefore limited visibility of the contribution of the compressors to the operation and risks of the system.

At the start of this case, in 2009, an internal company review considered how to align the operating philosophies of the different operating units; during this process the potential benefits of 'horizontal asset management' planning were examined. These were anticipated to be (1) improved productivity and flexibility, with employees and contractors using familiar work processes across the assets under management and (2) centralizing engineering and network design to minimize the duplication of design work that can be applied to similar assets.

45.4.2 Identification of Business Drivers

In moving gas from supply point to demand point (gas transmission) safely and efficiently the business drivers are Safety, Regulatory compliance, Environmental, Economic, and Customer Service. The level of service requirements for asset management relate directly to these business drivers.

- Safety—To maintain and operate assets to the extent that the risks to employees, contractors, and the public are maintained at as low as reasonably practicable.
- Regulatory Compliance—To meet all regulatory requirements associated with the gas Distribution License.
- Environmental—To maintain and operate assets so that the risks to the environment are kept at as low as reasonably practicable.

- Economic—To ensure that costs are prudent, efficient, consistent with accepted industry practises, and necessary to achieve the lowest sustainable cost of providing gas distribution services.
- Customer Service—To maintain and operate assets consistent with providing a high level of service (safety and security of supply) to consumers.

45.4.3 The Level of Service Table

The results of the LOS development for a compressor station and a single centrifugal compressor are presented in Tables 45.1 and 45.2. Minor changes and data omissions have been made to the table to protect Company sensitive information.

The format of the LOS table is to identify the business drivers in the left column and then consider what the asset needs to do to deliver on that level of service. The second column identifies the measure used to assess its performance, and the third column the target. Although not shown here due to confidentiality considerations, it is common to have two additional columns, one for actual performance and the final one to show, often using colors, if the performance is within, close to, or out of specification. In the Performance, Cost, and Condition section that follows the LOS table in the AM Plan, further details on the actual versus target performance is usually explored with the use of bar charts or other graphing approaches. These

Table 45.1 Level of service table for a gas transmission pipeline

Business drivers	LOS description	Performance measure	Example performance target
Safety	Safety	Total recordable injury frequency rate TRIFR	0
Regulatory compliance	Regulatory	Statutory inspections on the pipeline completed/scheduled (%)	100 %
		Non-compliance orders issued on the pipeline	0
Environmental	Emissions	CO_2 emissions per pipeline (MJ/total gas delivered)	<3 %
Economic	Cost-effectiveness	Pipeline maintenance costs ($/total gas delivered)	–
		Fuel consumption (m^3/day)	–
		Condition surveys scheduled/completed (%)	100 %
Customer service	Contractual requirements	System availability (%)	99 %
		Loss of supply events	0
		Transmission pipeline capacity performance	–

Table 45.2 Level of service table for individual compressor in a compressor station

Business drivers	LOS description	Performance measure	Example performance target
Safety	Safety	Total recordable injury frequency rate TRIFR	0
Economic	Cost-effectiveness	Compressor maintenance costs $/utilized hour	–
		Compressor scheduled maintenance work ratio (%)	80 %
		Utilization of the compressor (%)	70 %
		Compressor condition	–
Customer service	Contractual requirements	Availability of the compressor (%)	95 %
		Reliability of the compressor (MTBF)	–

graphs show historical performance and trends and can also be used to compare performance across the asset class.

45.5 Discussion of Results

45.5.1 Selection of LOS and Performance Measures

The selection of LOS and a discussion on the performance measures is provided below.

Safety LOS: 'Total Recordable Injury Frequency Rate' (TRIFR) is a common industry lagging measures. TRIFR records the rate of total recordable injuries respectively, per every 1 million man hours worked. TRIFR comprises all types of injuries, including: Lost time injuries; medically treatable injuries; and restricted work injuries. TRIFR can be measured at the pipeline and compressor level.

Regulatory LOS: The measures selected for a transmission pipeline for regulatory compliance include leading and lagging measures relating to getting inspections done and managing any excursions to ensure no non-compliance notices are received. The current focus on carbon emissions resulted in the inclusion of a measure on CO_2 emissions.

Environmental LOS: Currently there are no government requirements for emissions tracking however this may well change given developments in the climate change debate within Australia in 2011. It is possible that a measure of emissions will be required. It is shown here for a pipeline. There would be significant challenges in trying to measure this on a single compressor and it is not clear what the benefits would be.

Cost-effectiveness LOS: There are a number of measures that can be used to define how costs are managed and value is achieved. Ideally there would be a small set of measures that included leading and lagging measures. One lagging measure is maintenance cost, this can be determined at the pipeline (normalized using gas moved) and compressor levels (per unit). A ratio is required to allow for comparison across pipelines and compressors respectively. There are a number of leading measures of maintenance costs; 'scheduled maintenance work ratio' (actual hours on scheduled work/total hours) is widely used and is suggested here for the compressor. Scheduled maintenance work ratio measures the level of unscheduled work and level of control over the work management process. Fuel consumption is another cost and influenced by compression capability; this is measured for a pipeline. Fuel gas usage is not easy to measure at the compressor level.

Other measures for the compressors performance are 'availability', utilization and 'mean time between failure'. Availability is impacted by scheduled and unscheduled downtime of the individual units. Utilization allows managers to determine how the assets within the asset base are being utilized, are some over or under utilized? This is especially important in redundant systems like compressor stations. 'Mean time between failure' is a measure of the reliability of the unit.

Condition was widely regarded as an important leading measure of costs and reliability but there are significant challenges in determining the measure and approach, this is described further in Sect. 45.5.4 below.

Contractual requirements LOS: The Company exists to transport gas to its customers under contract; these contractual requirements include clauses around loss of supply, and availability of supply. The pipelines exist to move the gas through the system and hence two lagging measures of the pipeline performance are system availability and loss of supply events. Due to redundancy configurations at compressor stations and storage capacity in the transmission lines it is highly unlikely that failure of a single compressor will impact either of these two measures so there is no contractual measure in the LOS table for a single compressor. Another measure for the transmission pipeline is capacity performance, this identifies capacity constraints.

45.5.2 Examining Performance Against Targets

An example of how the performance of the major asset classes, transmission pipelines, gas storage, power generation, and compressors can be displayed for management is shown in Fig. 45.2. This simple display shows where there is deviation from targets in each of the LOS categories of safety, regulatory compliance, environmental, economic/cost effectiveness, and contractual requirements/customer service. The data shown is for illustration only and does not represent the company's actual performance.

There were a number of challenges with deciding which measure or set of measures to use for each LOS and how to normalize the measures to allow for comparison and trending.

LOS Summary	Safety	Regulatory compliance	Environmental	Economic/ cost-effectivess	Customer service
Pipelines					
Pipeline 1	OK	LOW	OK	OK	OK
Pipeline 2	OK	OK	OK	OK	OK
Pipeline 3	OK	OK	OK	LOW	OK
Pipeline 4	OK	OK	OK	OK	OK
Pipeline 5	OK	OK	OK	OK	OK
Pipeline 6	OK	LOW	LOW	OK	OK
Gas Storage					
Gas Storage 1	OK	OK	OK	LOW	OK
Power Generation					
Power Gen 1	OK	OK	OK	OK	OK
Power Gen 2	OK	OK	OK	OK	OK
Asset Classes					
Compressors	OK	n/a	n/a	LOW	OK

Fig. 45.2 Display of LOS measures for different asset systems (the data is for illustration)

- Each measure had to be achievable in the current state of the different business systems and operating teams.
- At the start of the project there was no standard definition of time across the business. Each engineer appeared to have their own way of determining measures such as availability and reliability. This was subsequently resolved as part of this project by developing a Company-specific 'Standard Definition of Time' based on the work of Katukoori [25].
- Deciding if the measure should be at the asset level, or at the pipeline level. For example, the availability of a single compressor at a station is not a direct indication as to if the contractual requirement is met. The key is to have a blend of appropriate measures.
- Recognition that measures drive behavior. Education is required to ensure that users of the metrics do not look at measures in isolation but consider a suite of measures to support informed decision-making and action.

45.5.3 Measure for Compressor Condition

Ideally there would be a measure to inform management about the "asset health"; this could be used to trend if the average health of the asset class and also individual health of at-risk compressors is stable or deteriorating. Within the case study Company there are condition assessments, integrity reviews, and other condition related activities. Sometimes this information is held by different people and in different systems but there was no process of integrating this to present an overall view of health or condition-related risk. A significant time in this project was exploring how this could be achieved, but it remains an area where future work is needed. The authors are aware of the work on-going research work at CIEAM and IMS to present an integrated view of asset health and there is work in the civil infrastructure sector. However, there was limited work the authors could draw on for a “one number” view of a health of a complex rotating asset, the compressors. As an interim measure a simple approach was proposed based on a 4-tiered Condition Index, representing the sum of the criticality ratings (normal, monitor, caution, severe) of a quantitative review of the following inspections: Vibration Analysis, Lube Oil Analysis, Boroscope Inspections, Temperature Trending Analysis, and Efficiency Analysis.

The aim of the index is to allow for comparisons between compressors regardless of operating condition and location, to aid in forecasting and budgeting. The condition index is presented in a 4-tier graph, as shown in Fig. 45.3. This figure provides a simple graphical tool to see is the condition of the asset class is improving or not. While the graph is an immediately useful tool, condensing the graph results into a single measure is a concern and so the condition index is not included in the LOS table at present. The details of how the index inputs are calculated may be the subject of a future paper.

45.5.4 Using LOS as an Alignment Process

“Most organizations attempt to create synergy, but in a fragmented, uncoordinated way. They do not view alignment as a management process” [2]. The approach described in Sect. 45.6 shows how the LOS can be used to describe a set of performance measures for compressors across the business, regardless of which asset system (vertical slices) it belongs to. Having established this, there can be an informed dialog with the different functional groups about the risks of failing to meet LOS objectives, and how to best prioritise resources and funds. This focus of how each functional group will achieve the LOS can be tailored to their specific skills and resources but the company can remain sure that all teams will maintain the same focus (by having exactly the same LOS to work towards) and that these LOS are directly related to corporate goals and demands.

Kaplan and Norton’s book ‘Alignment’ proposes that the annual planning process provides an architecture around which the alignment process can be executed. The annual review of AM Plans is an obvious vehicle in the asset management context. The book proposes eight alignment checkpoints for corporate, business units, and support units of typical multi-business organizations [2]. The process described above supports assessment of these alignment checks. In addition these provide the opportunity at the asset level to determine if there are

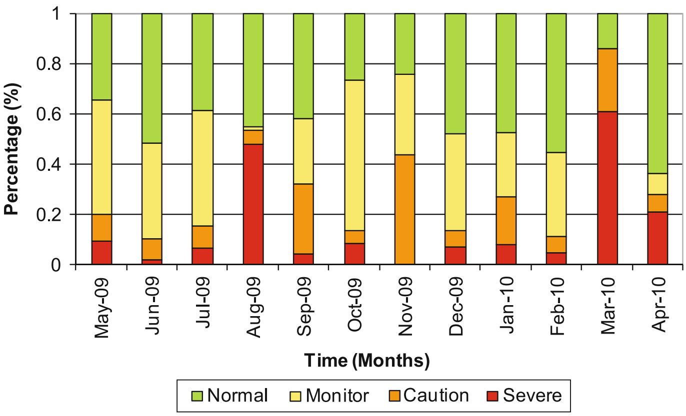


Fig. 45.3 Condition index for the compressor population across the company (example only)

opportunities to improve. Graphs showing the relative performance of assets, using common measures, across the business can be used to identify underperforming units. Conversely those responsible for high performing units can be encouraged to share best practice.

45.6 Conclusions

Organizational objectives need to be transformed from descriptive statements of desired outcomes for safety, regulatory compliance, environmental, economic, and customer service objectives to quantitative performance measures that managers of assets can use to guide decision making and assess performance. The Level of Service approach described in this paper supports this translation.

The case study shows how the Level of Service approach can be applied to a gas transmission network, looking at the asset system at two levels, a single transmission pipeline, and a compressor. Examples for how LOS performance measures can be selected and some of the challenges in choosing measures discussed. Of particular interest is determining how to set an asset health indicator for the compressors.

The case study shows how adoption of the LOS framework in a formal and centralized manner has the potential to reduce the likelihood of misalignment in interpretation of company objectives. These Level of Service measures also provide an effective communication platform between hierarchy levels from senior management to operational management and also functional groups such as engineering and accounting. Implementation of the approach was achieved by incorporating level of service summaries into the existing reporting structure to minimize the disruption and required organizational change.

The work highlights some of the many challenges organizations face as they attempt to create a line-of-sight view of their assets from the boardroom to the shop floor. It presents the outcomes of a process to create the alignment necessary to ensure that asset risks are managed and that the assets continue to deliver value to the operator.

References

1. Kaplan RS, Norton DP (2005) The balanced scorecard: measures that drive performance. Harvard Bus Rev 83:172–180
2. Kaplan RS, Norton DP (2006) Alignment—using the balanced scorecard to create corporate synergies. Harvard Business School Press, Boston
3. BSI (2008) Asset Management Part 1: Specification for the optimised management of physical infrastructure assets. British Standards Institute, England
4. Scott J (2006) BSI-PAS 55 certification and the stakeholder working group. The Office of Gas and Electricity Market, London
5. Jones A (2009) Asset management and regulation at energy safe victoria. In: Proceedings of ICOMS asset management conference. Sydney, Australia
6. Kaplan RS, Norton DP (1006) The balanced scorecard. Harvard Business School Press, Boston
7. Halfawy MR (2008) Integration of municipal infrastructure asset management processes: challenges and solutions. J Comput Eng 22(3):216
8. Maunsell (2002) International Infrastructure Management Manual, 2nd edn. Association of Local Government Engineering, New Zealand, Incorporated and National Asset Management Steering Group
9. Orion (2007) Orion asset management plan. Orion, Christchurch
10. Transportation Research Board (1985) Special report 209: highway capacity manual. Transportation Research Board, Washington DC
11. Chasey AD, de la Garza JM, Drew DR (1997) Comprehensive level of service: needed approach for civil infrastructure systems. J Infrastruct Syst 3(4):143–153
12. Moselhi O, Zayed T, Khan Z, Salman A (2010) Community-driven and reliability-based budget allocation for water networks. In: Proceedings of the 2010 construction research congress: innovation for reshaping construction practice, pp 578–587
13. Correia AR, Wirasinghe SC, de Barros AG (2008) Overall level of service measures for airport passengers terminals. Transp Res Part A: Policy Pract 42(2):330–346
14. Franques JT, Townsend RL (1998) Storm water level of service. In: Proceedings of the ASCE wetlands engineering river restoration conference. Denver, Colorado
15. Flannery A, McLeod D, Pederson NJ (2006) Customer-based measures of level of service. Inst Transp Eng J 76(5):17–21
16. Kononov J, Allery B (2003) Level of service of safety: conceptual blueprint and analytical framework. Transp Res Rec 1840:57–66
17. ASME (1997) Performance test code on compressors and exhausters. In: ASME PTC 10-19971998. American Society of Mechanical Engineers, New York
18. ISO 13631:2002 (2002) Petroleum and natural gas industries—Packages reciprocating gas compressors. International Organisation for Standardisation
19. ISO 14224:2006 (2006) Petroleum and natural gas industries—Collection and exchanges of reliability and maintenance data for equipment. International Organisation for Standardisation
20. OREDA (2002) Offshore reliability data handbook. SINTEF Industrial Management, Norway

21. Walsh P, Fletcher P (2004) Gas turbine performance, 2nd edn. Blackwell Science and ASME Press
22. Brun K, MG (2006) Nored, Guideline for field testing of gas turbines and centrifugal compressor performance. Southwest Research Institute
23. McKee R (1998) Continuous compressor performance monitoring: leveraging current technology for optimum benefits. In: Proceedings of the gas machinery research council gas machinery conference. Denver, Colorado
24. Neely A (2002) Strategy and performance: getting the measure of your business. Cambridge University Press, Cambridge
25. Katukoori VK (2007) Standardizing availability definition. University of New Orleans, New Orleans, pp 1–21

Chapter 46
Condition Assessment of Civil Engineering Assets

Joe Amadi-Echendu and Nkululeko Xulu

Abstract The management of engineered assets like cooling tower structures constructed from reinforced concrete, and pipe racks fabricated in steel, generally receive less maintenance emphasis than plant equipment such as instrumentation, pumps, switch gear or vehicles. This seems more apparent in mining and mineral processing firms where production output is the dominant concern. Coupled with the usually longer life span of civil structures and static mechanical components, this tendency often leads to limited knowledge about the condition of assets like cooling towers, substation buildings, and pipe racks which, in principle, cannot run to failure. The paper is based on an ongoing case study of a large industrial complex. Three sources of data are being explored to provide insight regarding the condition assessment of engineering structures in practice. The perplexing initial findings suggest that despite advances in condition monitoring technologies, more often than not, visual inspection prevails as the assessment method and thus confines budget provisions for maintenance activities.

46.1 Introduction

Engineered assets comprising our built environment typically consist of civil support structures and static mechanical components like columns, foundations, and vessels, electromechanical items like motors, as well as electronic instrumentation and computer systems. In a petrochemical complex, an appreciable number of civil engineering structures and static mechanical devices are installed. These static components generally form the basic infrastructure of plants and facilities that make up the asset base for the associated business concern.

J. Amadi-Echendu (✉) · N. Xulu
Graduate School of Technology Management, University of Pretoria,
Republic of South Africa
e-mail: joe.amadi-echendu@up.ac.za

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_46,

Such basic infrastructure is subject to weathering effects as well as subjected to high temperatures, leaks, and discharges of aggressive chemicals from process lines, and vibrations induced by dynamic equipment. These effects impact on the technical integrity of civil engineering structures (like reinforced concrete and structural steel) and influence their performance. The pertinent consideration is that such assets may deteriorate in time but should not fail to provide the intended function or service. Hence, the technical integrity of structural assets like steel pipe racks, cooling towers, buildings, and equipment supports must be continuously managed within their service life. Here, we define technical integrity as the "assurance that, within limits of *predictable* operating conditions, there is no foreseeable chance of failure that will endanger the safety of people, damage the environment, or adversely affect the business value of a *structural* asset…". The primary business goal is to prevent unexpected failure, which may result in loss of life and/or major financial adversity.

In principle, the design of a structure must satisfy two limit states:

1. Structural resilience (considers all factors that may lead to structural collapse—e.g. bending, shear, and tensile or compressive stresses), and
2. Serviceability (which relates to the functionality of a structure).
3. The serviceability may be further described in terms of performance criteria that include [1]: fitness for purpose, safety and reliability, durability, value for money, appearance, user comfort, and robustness.

In practice, poor design, construction, and workmanship in conjunction with aggressive industrial and natural conditions such as atmospheric and ground pollution all contribute toward structural degradation and subsequent deterioration in desired performance of the asset. We can take advantage of advances in sensor and sensing technologies [2] to monitor structural degradation through non-destructive inspection tests. Data obtained from such tests may be combined with information extracted from other sources to provide indication of the level of deterioration or condition of the asset.

This paper discusses a case study of a large industrial complex. We explore three sources of data and information in order to provide insight regarding the practice of condition assessment of civil and mechanical engineering structures. The study assumes that relevant legislations regarding structural assets are applied at the case study environment. Section 46.2 includes a very brief summary on failure modes and proposed indicators for condition assessment of civil engineering structures. In Sect. 46.3, we describe the sources of data and information and the findings. The paper concludes with some inferences based on the findings.

46.2 Failure Modes and Condition Indicators

A significant proportion of infrastructure assets are constructed using reinforced concrete and steel such that defects and failure modes of these materials provide indication of degradation and deterioration of the assets. For instance, on the one hand, reinforced concrete is a material of choice for assets like cooling towers, culverts, and substation buildings. Like any other construction material, it is susceptible to distress and deterioration as a result of accidental loadings, chemical reactions, construction errors, design errors, movements, and temperature changes. The typical failure modes for reinforced concrete are cracking and disintegration and the extent of the defect provides some indication of the condition of the asset. According to Winterbottom and Goodwin [3], the most common causes of cracks in concrete are chemical hardening, expansive reaction loads, thermal stresses, and shrinkage. Woodson [4] valuable guidance on the measurements crack patterns, width, and depth. He suggests that crack width may be classified as fine (i.e., less than 1 mm), medium (1–2 mm), and wide (excess of 2 mm). He also advices that crack depths may be categorized as *surface*, *shallow*, *deep*, and *through*.

On the other hand, steel is favored for the construction of items like conveyor supports, pipe racks and tanks. For assets fabricated using steel, the prevalent failure mode arises from material corrosion caused by exposure to aggressive natural and man-made chemicals. Another latent but common failure mode is fatigue, typically induced by overloading stress that is readily propagated through imperfections in the material. The identification of corrosion rate through wall thickness measurements is a well established art. Programmed inspection of assets constructed from reinforced concrete or fabricated from steel via visual means and/or non-destructive testing techniques can reveal progressive changes in failure modes, so that appropriate mitigation measures may be taken to reduce degradation, damage or deterioration and to avoid or prevent failure. To state what the condition of the asset is, we propose that the extent of degradation, damage or deterioration must be compared to the capacity of service required. In this regard, and extrapolating from Refs. [5–7], we illustrate generic descriptors for condition assessment as in Table 46.1.

For our case study environment, we adapted the generic descriptors for reinforced concrete and steel materials as shown in Table 46.2.

46.3 Research Findings

Having established the preceding framework and model for assessing the condition of structural assets constructed from reinforced concrete and fabricated from steel materials, we then collected data on cooling towers, substation buildings, culverts, and pipe racks installed in the petrochemical complex used as our case study environment. Our approach involved discussions with practitioners, collecting records of inspection data, and site visits to confirm the existence of the assets. Our discussions with practitioners indicated that the method most applied was visual inspection, even though they are aware of other non-destructive techniques (NDT) widely available. From the discussions, it became apparent that cost, sophistication, and analytical complexity increased the levels of trepidation in the use of NDT hardware and software.

Table 46.1 Generic descriptions for condition assessment of an engineered asset

	Generic description	Condition code	Condition index
Optimal performance	Fully loaded under optimal operating environment, probability of failure <10 %, PF interval >0.7, compliant to standards and legislation, no maintenance backlog	A = Very Good	Greater than 80
Normal operating conditions	Optimal loading, probability of failure >30 %, PF interval <0.7, compliant to legislation and standards, minimum maintenance backlog	B = Good	60–80
Regulation compliance	Reduced loading, failure probability >50 %, PF interval <0.5, compliant to legislation and many standards, maintenance backlog = planning horizon	C = Serviceable	40–59
Vulnerability to conditions	Reduced loading, probability of failure >70 %, PF interval <0.3, compliant to legislation, maintenance backlog > planning horizon	D = Critical	20–39
	Minimum loading, probability of failure >80 %, PF interval <0.2, non compliant	F = Unserviceable	Less than 20

Table 46.2 Condition assessment descriptors for reinforced concrete and steel materials

Failure mode		Condition code	Condition index
Reinforced concrete	Steel		
Cracks (>2 mm; wide, and through)	Section loss due to severe corrosion	Unserviceable	≤10
Spalling resulting in section loss	Major deflection	Unserviceable	≤20
1 mm < crack < 2 mm (medium and deep)	Paint loss and freckled rust	Useable	≤40
Minimal spalling resulting in exposure	Medium deflection	Useable	≤40
Cracks (<1 mm; fine, and shallow)	Minimal surface corrosion	Good	≤60
Surface abrasion	Negligible deflection	Good	≤70
Negligible cracks (≪1 mm; fine, and surface)	Negligible corrosion or deflection	Very good	≥70

Table 46.3 Example of a visual inspection record

Asset type	Substation building	Asset ID	Inspector
Materials	Brickwalled reinforced concrete and asbestos roofing	5,500 J	A N other
Date of inspection	3 September 2009		
Weather condition	Clear sky and warm temperature		

Subcomponent	Failure mode	Condition code	Index
Floor slab	Smooth finish with fine and shallow cracks	Good	65
Walls	Collapsed wall showing exposed switch panels	Unserviceable	15
Roof	Broken panel and beams showing signs of corrosion	Unserviceable	15
Recommendation	Urgent revamp work necessary	Critical	

The dilemma is that visual inspection cannot detect cracks within the material and this begs to question the validity of data recorded. A sample visual inspection record is illustrated in Table 46.3.

Visual inspection records were obtained for seven cooling towers of the same age and similar operational environment but differing maintenance history. Our interpretation of the visual inspection data using the condition assessment descriptors suggests that two towers appeared to be in 'good' condition, four towers in 'useable' condition, and one tower was in a 'critical' condition. Similar data was collected for seven randomly selected substation buildings, six different pipe racks, five culverts and one bridge. The results also suggest that:

1. Three of the older substation buildings were becoming 'unserviceable', three were in 'serviceable' condition, while the latest two were in 'good' condition;
2. Two pipe rack were in 'good' condition, three were 'useable' and one 'unserviceable'; The bridge and three culverts were only 'useable' while two culverts were 'unserviceable';

3. Overall, substations, and pipe racks seemed to be in better condition that cooling towers and culverts.

Although it is very limiting and subjective, visual inspection data can be used for initial condition assessment to guide budget estimation and, the preliminary results presented here are useful in that sense. The ongoing study is suffering from reluctance to release information and confidential records of NDT inspections. This limits the application of the framework and model briefly described in Sect. 46.2 of the paper.

References

1. Ray SS (1995) Reinforced concrete: analysis and design. Blackwell Science Ltd, UK
2. von Holdt CJ, Belmonte HMS, Amadi-Echendu J (2009) Review of technology used in strategic asset management: existing and future needs. Water Research Commission, South Africa. Report No.: 1785/1/09
3. Winterbottom G, Goodwin F (2005) Concrete cracks: causes, correcting and coating. J Prot Coat Linings 22(1):40–52
4. Woodson RD (2009) Concrete structures protection, repair and rehabilitation. Heinemann, Butterworth
5. Amadi-Echendu J, Belmonte HMS, von Holdt CJ, Bhagwan J (2009) A case study on condition assessment of water and sanitation infrastructure. In: Proceedings of WCEAM Athens Greece, 28–30 Sept 2009, p 173
6. Rahman S, Zayed T (2009) Condition assessment of water treatment plant components. J Perf Constr Fac 23(4):276–287
7. Mishalani RG, McCord MR (2006) Infrastructure condition assessment, deterioration modeling, and maintenance decision making: methodological advances and practical considerations. J Infrastruct Syst 12(3):145–146

Chapter 47
Analysis of Wear Mechanisms in Pneumatic Conveying Pipelines of Fly Ash

A. A. Cenna, K. C. Williams, M. G. Jones and W. Robinson

Abstract Pneumatic conveying is a frequently used method of material transport particularly for in-plant transport over relatively short distances. This is primarily to exploit the degree of flexibility it offers in terms of pipeline routing as well as dust minimization. Approximately 80 % of industrial systems are traditionally dilute phase system which uses relatively large amount of air to achieve high particle velocities to stay away from trouble, such as blocking the pipeline. However, for many applications higher velocities lead to excessive levels wear of pipelines, bends, and fittings. To combat these problems, many innovative bends have been designed. These designs have solved the problem of wear in the bends, but often introduce the wear problem in the area immediately after the bend due to the changed flow conditions. Wear in pneumatic conveying is a very complex problem and at present there is limited understanding of the wear mechanisms responsible for the severe wear in certain areas of a pneumatic conveying pipeline. The ability to determine the wear mechanisms in these areas holds the key for determining the service life of pneumatic conveying pipelines in industry. Even though the fly can be conveyed at low velocity dense phase mode, wear of pipeline conveying fly ash remained a critical issue for many power plant operators. In this paper the wear mechanisms in a fly ash conveying pipeline has been analyzed. Wear samples from fly ash conveying pipeline have been collected and analyzed for dominant wear mechanisms in the critical wear areas. Analysis of the worn pipeline showed continuous wear channels along the bottom of the pipeline consistent with the abrasive wear by larger particles. The other severe wear areas are the sections after the special bends used to reduce bend wear. Scanning electron microscope (SEM) analysis of the surfaces revealed that both erosive wear and

A. A. Cenna (✉) · K. C. Williams · M. G. Jones
Faculty of Engineering and Built Environment, Centre for Bulk Solids and Particulate Technologies, University of Newcastle, Callaghan NSW 2308, Australia
e-mail: Ahmed.Cenna@newcastle.edu.au

W. Robinson
Delta Electricity, NSW Power Generations, Wallerawang, NSW, Australia

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_47,

abrasive wear mechanisms are present in these areas. Formation of a surface layer similar to transfer film in alumina conveying pipelines have been recognized in this analysis. These layers seem to be removed through brittle manners such as cracking and spalling. The wear mechanisms and the wear debris seen on the surface are consistent with wear by larger particles.

47.1 Introduction

Pneumatic conveying involves the transportation of a wide variety of dry powdered and granular solids through pipeline and bends using high pressure gas. It is a frequently used method of material transport particularly for in-plant transport over relatively short distances. This is primarily to exploit the degree of flexibility it offers in terms of pipeline routing as well as dust minimization. Approximately 80 % of industrial systems are traditionally dilute phase system which uses relatively large amount of air to achieve high particle velocities to stay away from trouble, such as blocking the pipeline. However, for many applications higher velocities lead to excessive levels of particle attrition or wear of pipelines, bends, and fittings. To combat these problems, there are systems designed to operate at relatively low velocity regimes. Yet one problem remains as a major issue with these conveying systems, which is wear.

Silicon dioxide (SiO_2) or silica and aluminium oxide (Al_2O_3) or alumina are the two major components in the chemical composition of fly ash. The percentage of silica can be as high as 65 % and alumina can vary between 15 and 30 %. Both alumina and silica are very hard materials, with silica having a hardness value of about 6 on the mohs scale of hardness and that of alumina being close to 8 [1]. Due to the high concentration of these two hard constituents, fly ash is very abrasive and cause serious damage to the pipeline surfaces either through abrasion of impact.

Fly ash is the major by-product of coal fired power stations. Ash is collected from along the path of the flu gas. Based on the collection route of fly ash, particle size varies which affects the pneumatic conveying of fly ash considerably. The ash is first collected in economizer hoppers, and then air pre-heater hoppers before the series of electrostatic precipitator (ESP) hoppers. About 85 % of the total ash carried with the flue gas is collected in the ESP hoppers. The average particle size of the ash collected from the economizer hoppers is about 125 μm, whereas particle collected from ESP is much finer. Typical particle size of ash based on the collection points is given in Fig. 47.1.

Although there is no difference in the constituents of the fly ash collected from different points, particle size distribution have a significant influence on the conveying capability of the material. The majority of the fly ash particles are expected to be spherical glassy droplets as a result of the combustion process. In fact the

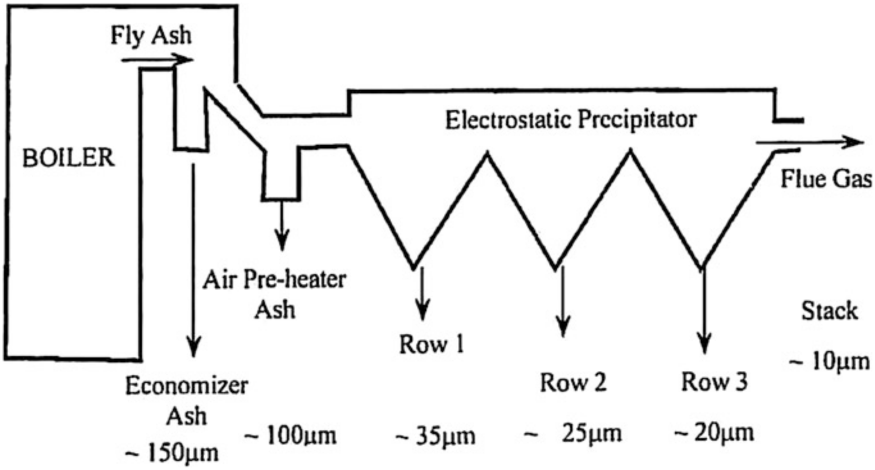


Fig. 47.1 Typical particle sizes of ash based on the collection point in a coal fired power station [1]

particle shapes and sizes vary considerably as can be seen from the SEM images (Fig. 47.6b). Quality of coal used in the combustion process might have an effect on the shape of ash particles.

Wear mechanisms in the pipeline depends on the types of particle interactions with the pipeline surfaces. It is an accepted knowledge that the wear in pneumatic conveying pipelines is primarily erosive in nature. As a result many innovative techniques and bends have been developed over the years to combat a particular type of wear, erosion. In many cases these innovative bends have successfully saved the bends from erosive wear but often introduced new wear area depending on the flow characteristics in the bend. One such situation has been analyzed in this paper.

47.2 Wear in Industrial Pipelines

It is well-known that the bends are primarily highly vulnerable to wear in pipeline conveying abrasive materials like fly ash and alumina. In many cases specially designed bends are used to combat the wear situation in bends. In particular, concrete filled short radius bends are used to protect the bends from wear in this industry. Using concrete filled short radius bends the wear in the bend area has been controlled but the problem is not over. The wear problem has just been relocated from the bends to the section immediately after the bend. The pipeline in these areas (section after the bend) reportedly has a nominal life expectancy of 4–6 months only.

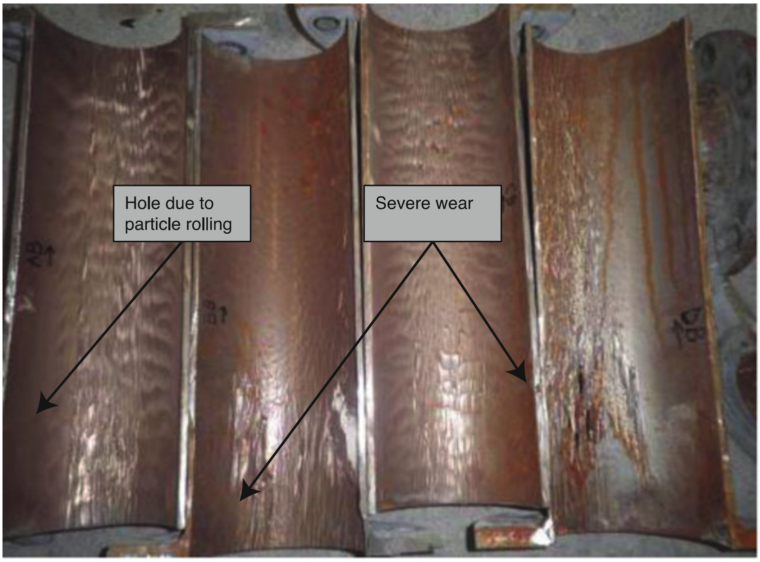


Fig. 47.2 Wear patterns in different areas critical with respect to service life of fly ash conveying pipelines

To analyze the wear situation in the pipeline, wear pipes from different locations have been collected for analysis. Figure 47.2 shows wear sections from different location of the conveying pipeline. Not all the sections are critical with respect to the conveying pipeline; as such they do not limit the service life of the pipeline within the same time period. Section A (AB refers to bottom part of section A) was taken from the middle of a straight section of the pipeline away from any bend effects. It shows the wave pattern on the pipeline representing the erosive wear by fine particle like fly ash. The wear patterns on the bottom of the pipeline are due to the larger particles sliding and rolling freely along the bottom of the pipeline. The wear grooves are clearly representing the sliding-abrasive wear due to particles sliding on the bottom surface. Rolling and sliding of larger particles along the bottom generate fatigue damage on the surface which accumulates to cause unpredictable failure of the pipeline. Such a hole on the pipeline is shown in Fig. 47.2.

Section D and B are particularly service life limiting sections. Section B has been taken from the pipeline after a concrete filled bend and section D was taken from the end of a shallow angle downward section. This is the start of the straight section. As the particle slowed down at the beginning of the straight section, the area is in particular subjected to abrasive wear superimposed by impact erosion of incoming particles. Distance of the wear area from the flange might be due to a misalignment of the pipes at the joint. Section B has been taken from the section immediately after concrete filled bend. This area (location B) has been one of the major concerns of the industry with respect to wear.

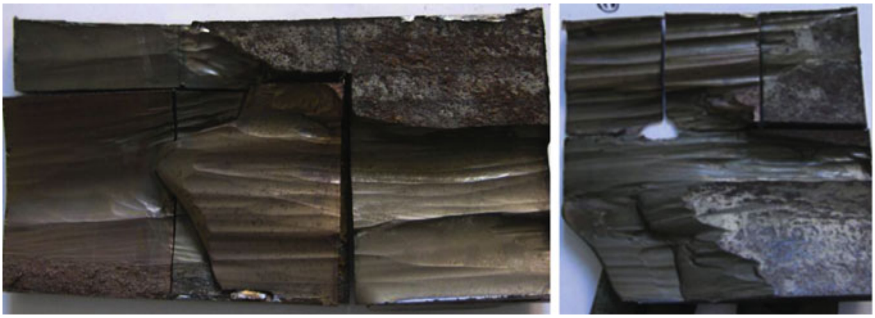


Fig. 47.3 Wear sections protected from corrosion and samples for SEM analysis

47.3 Wear Mechanisms in Critical Wear Areas

For a better understanding of the material removal mechanisms in the severe wear areas, wear samples have been collected for microscopic analysis. As seen in the Fig. 47.3, the samples are oxidized (rusted) very quickly after removal from the operations. This has been very critical to collect wear samples from the industrial pipeline and to protect them from oxidation before conducting the surface analysis. Finally the pipe sections were collected on the same day of replacement and protected using antirust before cutting the samples for SEM analysis. The wear samples have been preserved in a desiccator for SEM analysis. Figure 47.3 presented the pipe sections protected from corrosion and samples cut from for SEM analysis.

47.3.1 Erosive Wear by Fine Particles

The primary material removal mechanism in the critical wear areas is erosion. Visual examination clearly demonstrated this as well. Figure 47.4 presented the erosive wear mechanism observed in the fly ash conveying pipeline. Figure 47.6a presented a complex surface profile generated in the critical wear areas after the bend. This is due to the fact that these areas generate very high turbulence due to the particle impacts in the bend as well as reacceleration of particles after the bend. After the initial wear, particles tend to follow a preferential flow path within the particle flow structures. Due to particle segregation after the impact in the bend, larger particles tend to follow the bottom of the pipeline and majority of particles develop the core flow that impact on surface after the bend. Figure 47.4b presented some particle sticking to the surface even after ultrasonic cleaning of the surface. The variations in particle shape can be easily captured from this image.

Figure 47.4c clearly showing that the surface represented more cutting than deformation represented by shallow long grooves. This is different from usual

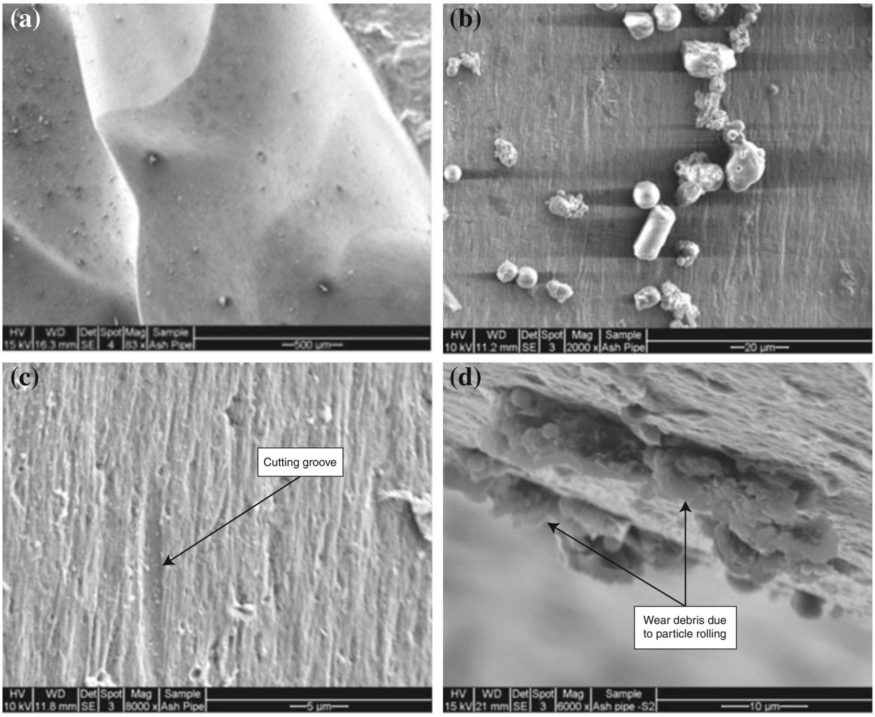


Fig. 47.4 Surface features representing the erosive wear mechanism

impact erosion where particle leaves the surface after impact whereas the particle kept in contact with the surface due to oncoming particles. Figure 47.4c showed an area where the impacted particle rolled on the surface and the wear debris is still attached to the surface. In fact, the surface features are more consistent with impact wear with added abrasive wear characteristics.

47.3.2 Abrasive Wear by Larger Particles

Surface characteristics consistent with abrasive wear are clearly visible in the bottom section of the straight section of the pipeline as shown in Fig. 47.2. Abrasive wear usually arises from the larger particles in the particle stream that cannot remain suspended at the conveying air velocity. These particles are free to roll or slide along the bottom of the pipeline. Larger particles accumulated in the bends as more particles dropped out of the air stream after the impact due loss of kinetic energy. These particles create the abrasive wear situation immediately after the bend as they lag behind the smaller particles being reaccelerated. Complex wear situation arises from the accumulated particles in the bend as the air velocity

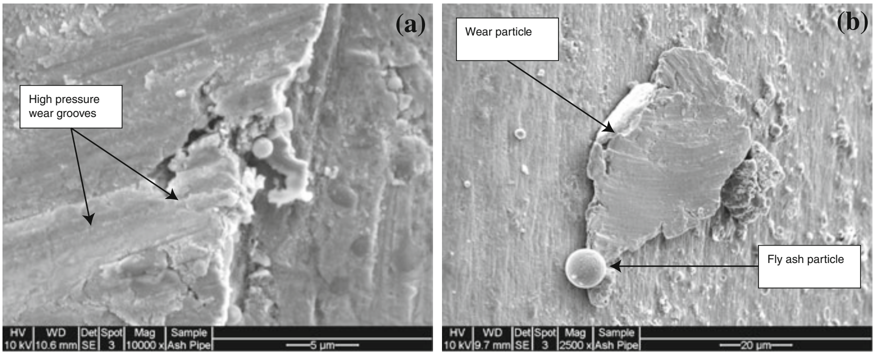


Fig. 47.5 Surface characteristics consistent with abrasive wear

due to reduced area of the pipeline. This helps to clear the pipeline faster with higher implication on wear as well. This process may increase the pressure fluctuation in the conveying pipeline.

Although the fly ash particles are expected to be spherical, particle attached to the surface (Fig. 47.4b) showed that they vary considerably. Figure 47.5 presented the surface characteristics consistent with the abrasive wear in the pipeline after bends. Figure 47.5a showed an area of interaction with a larger particle. Surface characteristics here are consistent with high pressure abrasive wear. Figure 47.5b showed a wear particle that has been detached from somewhere due to abrasive contact. A fly ash particle adjacent to the wear particle (Fig. 47.5b) bears the testimony of larger particle present in the fly ash.

47.3.3 Surface Modification and Formation of Surface Film

Detailed study of the formation of transfer film and modification in alumina conveying pipelines can be found in Glardon and Finnie [2]. One of the very unique characteristics of the wear surfaces observed in this study was the formation of transfer film and removal of surface layer through generation of surface cracking and crack propagation. These features are seen primarily along the bottom of the wear grooves in the severe wear areas in the straight section of pipeline after the bends.

Similar phenomenon has also been observed in the surfaces of the fly ash conveying pipelines. Fly ash is distinctively different from alumina both in shape and constituent materials. Although both the particles are extremely hard due to their constituents, alumina is highly angular while fly ash is mostly spherical particles. Both alumina and fly ash silica are ceramics. Although the formation of alumina transfer film has been understood to some extent, the formation of brittle layer observed in fly ash conveying pipelines is yet to be investigated.

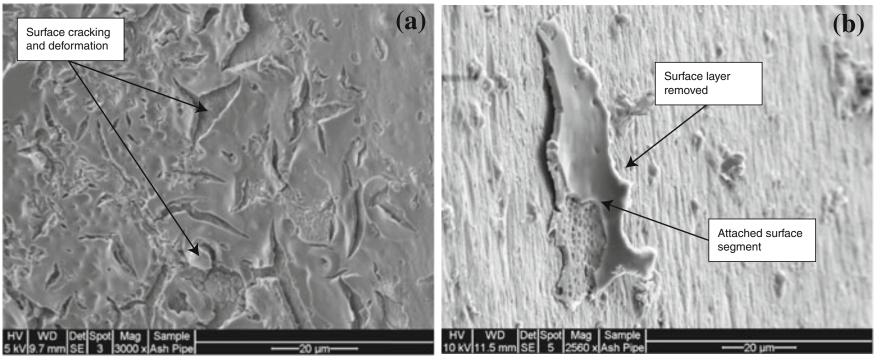


Fig. 47.6 Surface modification and material removal through brittle mechanism

Figure 47.6 presented an area from where the surrounding surface layer has been removed leaving a small segment still attached to the surface. The surface segment appears to be brittle in nature. Initial investigation of the surface layer showed that the surface was modified by the silicon, one of the primary constituent of fly ash.

The behavior of a hard layer on top of a relatively softer substrate had been explained through the characteristics of a thin hard coating on a softer substrate [3]. If the hard coating is not sufficiently supported by the substrate, it fails at relatively low contact stress because it cannot follow the deformation of the substrate (so called eggshell effect). Although the hard layer tends to protect the softer substrate from cutting and deformation wear, it fails due to the delamination and subsequent cracking and chipping of the hard layer. Cyclic loading and unloading can cause fracture or fatigue cracks, which ultimately destroy the harder surface layer [4]. The mechanism for delamination and cracking and subsequent removal through spalling has been presented for the case of alumina transfer film on mild steel pipeline by Cenna et al. [5].

47.4 Conclusion

From the observation of the industrial pipeline for wear patterns together with the wear mechanisms, it became clear that surface modification of conveying pipeline play an important role in determining the service life. The effects of the larger particle in the particle stream also a major factor for severe wear in the critical wear areas of the pipeline. Together with these two extraordinary effects, high velocity erosive wear is the primary factor for determining the service life of fly ash conveying pipelines.

It is possible that some carry over fly ash from the economiser hoppers are being mixed with ESP fly ash. This introduces larger particles in the fly ash convening system which seems to be a major contributor to the wear in the critical areas of the pipeline. As the larger particles always lag the finer grades of ash, these particles are usually cleared in the purging cycle at higher velocity. It would be reasonable to think that reduction of the larger particles in the system would increase the service life of pipelines considerably.

Acknowledgments This work forms part of the research program for the CRC for Integrated Engineering Asset Management.

References

1. Mills D, Jones M, Agarwal V (2004) Handbook of pneumatic conveying engineering. Marcel Dekker, Inc, New York
2. Glardon R, Finnie I (1981) A review of the recent literature on unlibricated sliding wear of dissimilar metals. J Eng Mater Technol 103:333–340
3. Meneve J, Vercammen K, Dekempeneer E, Smeets J (1997) Thin tribological coatings: magic or design? Surf Coat Technol 94–95:476–482
4. Holmberg K (1992) A concept for friction mechanisms of coated surfaces. Surf Coat Technol 56(1):1–10
5. Cenna AA, Williams KC, Jones MG, Page NW (2009) Generation of transfer film and its effects on wear mechanisms in alumina conveying pipeline of mild steel. Wear 267(1–4): 362–367

Chapter 48
RFID-Based Asset Management of Time and Temperature Sensitive Materials

M. Zawodniok, S. Jagannathan, C. Saygin and A. Soylemezoglu

Abstract RFID technology enables tracking of individual assets throughout the supply chain and real-time inventory management. However, the potential benefits cannot be realized until +99 % read rate is achieved. The radio frequency interference, signal fading, and harsh environment often result in not detected items, and subsequently incorrect asset management decisions. This paper presents a novel RFID-based smart freezer with a new inventory management scheme for extremely low temperature environments. The proposed solution utilizes a distributed inventory control (DIC) scheme, systematic selection of antenna configuration, and antenna power control in order to addresses those challenges. The performance of the DIC scheme is guaranteed using a Lyapunov-based analysis and verified in simulations and experiments on industrial freezer test bed operating at -100 °F. The proposed RFID antenna configuration design methodology coupled with locally asymptotically stable distributed power control (LASDPC) ensures a 99 % read rate of items while minimizing the required number of RFID antennas in the confined cold chain environments with non-RF friendly materials.

M. Zawodniok (✉) · S. Jagannathan · A. Soylemezoglu
Department of Computer and Electrical Engineering, Missouri University of Science and Technology, Rolla, Missouri 65409, USA
e-mail: mjzx9c@mst.edu

S. Jagannathan
e-mail: sarangap@mst.edu

A. Soylemezoglu
e-mail: soylemez@mst.edu

C. Saygin
Mechanical Engineering Department, University of Texas San Antonio, San Antonio, Texas 78249, USA
e-mail: can.saygin@utsa.edu

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_48,

48.1 Introduction

Inventory control of time and temperature sensitive materials (TTSM) are a vital problem that needs attention for various industries, for example composite manufacturers and food producers. Such items have to be stored in a temperature-controlled environment and used within a limited time, otherwise, they become waste. Effective management of TTSM will reduce inventory levels and prevent usage of expired materials thus reducing costs.

The goal of inventory management is to make the required quantities of items at the right time and location. Today, in inventory management, barcodes are widely used; however, this technology needs a line-of-sight is maintained for the items and also the items are scanned one at a time. In addition, barcode-based systems have problems under low temperature environments. Additionally, tight control of certain item types that are critical for production may be required. Consequently, tracking of all inventory items through traditional methods may not be economically feasible.

These limitations can be mitigated with RFID technology, where electronic tags programmed with unique identification code are attached to items. However, in order to provide item-level visibility through monitoring, tracking, and identification of items, a proper RFID implementation is necessary.

The application of RFID to inventory management was discussed in many papers [1, 2]. Those utilize the RFID technology with computational intelligence algorithms to improve efficiency of logistics control and management systems. Another application of RFID technology is localization of assets in warehouses and throughout a supply chain [3]. The availability of item-level visibility also provides opportunities to control another type of inventory, namely work in process inventory [4–6] and shelf inventory [7, 8]. In addition, RFID coupled with other sensing schemes helps close the loop on product life-cycle management (PLM) [1, 9] by providing visibility of product related information such as usage data, disposal conditions, etc. from cradle to grave.

However, the work in [1–13] assumes that a straightforward deployment of RFID technology providing almost 100 % read rates under "controlled" laboratory settings. Upon technology transfer, unforeseen problems, such as readability may arise from radio frequency (RF) interference, RF absorbing materials, and environmental conditions [14, 15]. For instance, in the proposed smart freezer, the low 100 °F below zero temperatures cause excessive ice buildup rendering RFID technology unreliable. In addition, RF absorbing materials render low read rates and longer reading times.

Therefore in this paper, a novel scheme of obtaining high read rates within a reasonable reading time for inventory management at extremely low temperature environments using passive RFID tags is introduced. This overall scheme consists of: (a) a backpressure-based inventory management scheme for reducing waste and to attain cost savings; (b) a selection scheme for placement and number of antennas inside the freezer; and (c) a locally asymptotically stable distributed power control (LASDPC) scheme or simply DPC scheme for antenna transmission in order to obtain a high read rate required for backpressure-based inventory management. Lyapunov approach [13, 14] is utilized to show the stability of the inventory control and the DPC schemes. The overall method has been experimentally tested and verified on hardware. Both simulation and experimental results are included to justify theoretical conclusions.

48.2 Methodology

In this paper, a novel distributed inventory control (DIC) scheme for balancing the inventory levels at each layer at a desired target baseline value using the backpressure mechanism along the supply chain is proposed. Figure 48.1 illustrates the supply chain structure with buffers, for example freezers on the shop floor, with local inventory control logic at each layer. This scheme considers only a single type of material flow. However, additional types of materials can be accommodated by using the same control logic for each material. The details of the DIC are presented in Subsection A under the assumption that there is full visibility.

In the subsequent section, antenna placement and power control discussion is presented in order to attain the 99–100 % read rate requirement. The process of selecting the number of antennas and their location is presented in Section C. Additionally, a novel mechanism that controls the power of the antennas to reduce the RF interference among them is proposed in Section II.D. The proposed antenna placement mechanism together with the DPC will be referred hereafter as the Missouri S&T scheme, which is shown to reduce the number of antennas inside the freezer, while guaranteeing 99 % read rates by eliminating all nulls.

The proposed DIC scheme expects a high item visibility. The proposed design methodology attains 99–100 % read rates by considering several factors:

- Number of antennas and their placements inside the freezer was altered to attain the RF coverage;
- The power on each antenna is varied using the proposed DPC, which reduces RF interferences and collisions in the backscatter direction.

The design methodology is given in Sects. 48.2.1 and 48.2.2.

48.2.1 Inventory Management Across the Supply Chain

48.2.1.1 Demand Prediction

This paper proposes a new model to predict the instantaneous demand given by

$$f(k) = [\alpha(k-1)\beta(k-1)]\begin{bmatrix} f(k-1) \\ D(k-1) \end{bmatrix} = \theta^T(k-1)\phi(k-1), \tag{48.1}$$

where $f(k)$ is the demand at time k, $\alpha(k-1)$ and $\beta(k-1)$ are online tunable parameters, and $D(k-1)$ is the actual measured demand at the time instant $(k-1)$. The error in prediction is then expressed as

$$e(k-1) = f(k-1) - D(k-2). \tag{48.2}$$

Selection of appropriate values for α and β is essential in order to improve the accuracy of the prediction model. Traditionally, these values are selected through offline analysis of historical data. In contrast, the proposed scheme performs online tuning of these parameters to guarantee stability as presented in Theorem 1.

Theorem 1 *Consider the instantaneous demand expressed as (1) with the error dynamics given in (2). Let the parameters update be provided by*

$$\theta(k) = \theta(k-1) + \gamma\phi(k-1)e(k-1), \tag{48.3}$$

where $0 < \gamma < 1$ *is the gain factor, then the instantaneous demand approaches the measured value asymptotically.*

Subsequently, the instantaneous value is utilized by the inventory controller, to create a backpressure signal, z, as shown in Fig. 48.1. The signal is propagated along the supply chain to adjust the delivery amount periodically. Next, the analysis of the backpressure mechanism is discussed.

48.2.1.2 Back Pressure Mechanism in the Supply Chain

In this section, inventory level dynamics in a supply chain are modeled using difference equations. In order to maintain target inventory levels, the DIC scheme is developed employing control theory. The performance of DIC is mathematically studied and the error bound on inventory level is obtained through the Lyapunov-based analysis.

At a particular echelon j in a supply chain, the inventory level will decrease when materials are used or shipped out to next echelon in the supply chain. When the change depends on the usage, it has to be estimated using (1). Otherwise, the change is dictated by the backpressure signal, $z_{i,\,j-1}$, which indicates the delivery request from the subsequent echelon, $j-1$. Then, the demand is fulfilled by delivery, *f*, shipped out from the current echelon of the supply chain. On the

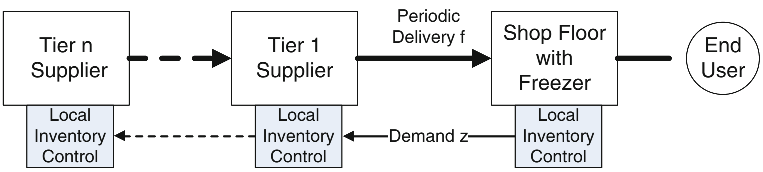


Fig. 48.1 Supply chain with back pressure mechanism

other hand, the inventory will increase when the material is delivered from the supplier up the supply chain based on demand *zij*. Thus in a supply chain environment, the inventory levels change due to the fluctuation in demand.

Overall, the change of the inventory level inside a freezer in the supply chain at echelon j can be expressed as

$$q_{ij}(k+1) = Sat_Q\left[q_{ij}(k) + z_{ij}(k) - f_{ij}(k) + e_{ij}(k)\right], \tag{48.4}$$

where $q_{ij}(k)$ is inventory level at freezer i at time k, $z_{ij}(k)$ is the required delivery for freezer i for period k, $f_{ij}(k)$ is the usage rate of the freezer I, and Sat_Q is saturation function modeling the limited capacity of freezers. The fluctuation in the demand, $e_{dij}(k)$, is due to many factors, for example just in time manufacturing, machine break down, etc. In general, the assets are shipped from a supplier toward the freezer/users. The reverse direction is also viable in the model through aggregation of the incoming flows.

Define the desired level of inventory as q_{tij}. Then the error in inventory level is equal to $e_{ij}(k) = q_{tij} - q_{ij}(k)$. The requested delivery level $z_{ij}(k)$ should minimize the error in inventory level for the next time instant. Unfortunately, the exact demand $f_{ij}(k)$ is unknown for the time period k (future) due to uncertainties such as usage rates, etc. By using an estimated demand value, backpressure control signal can be obtained as

$$z_{ij}(k) = \left[\hat{f}(k) + k_v e_{ij}(k)\right], \tag{48.5}$$

where $\hat{f}_{ij}(k)$ is the predicted demand for the freezer i during the time interval k in echelon j. The following theorem states that the error in inventory level is shown bounded, that is limited to a small error. The Theorem 2 is proven using Lyapunov analysis.

Theorem 2 (General Case) *Consider the desired inventory level,* q_{tij}, *to be finite, and the demand fluctuation bound,* e_{dM}, *to be equal to zero. Let the delivery quantity for (4) be given by (5) with the delivery quantity being estimated properly such that the approximation error* $\tilde{f}_{ij}(\cdot)$ *is bounded above by* f_M. *Then, the inventory level backpressure system is uniformly ultimately bounded provided* $0 < k_v < 1$.

48.2.1.3 Inventory Balancing in a Multi-Freezer Environment

In a typical industrial environment, there may be more than one freezer as shown in Fig. 48.2. Then, the overall inventory of the shop floor is expressed as a set of equations given by (4), where $i = 1, \ldots, n$. In such a case, the error in inventory at the shop floor, $e(k)$, is equal to sum of error magnitudes at each freezer,

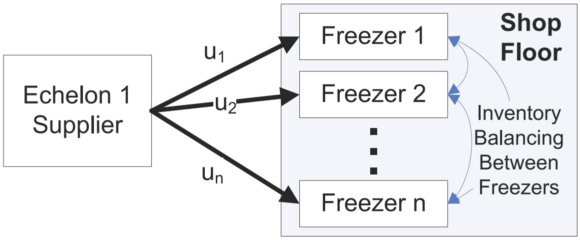


Fig. 48.2 Inventory balancing between multiple freezers

$$e_j(k) = \sum_i |e_{ij}(k)|.$$

In the multi-freezer scenario, as shown in Fig. 48.2, items can be quickly moved between the freezers to balance inventory, minimize wastage thus saving cost. Due to the co-location of freezers, the load balancing mechanism can yield almost immediate (relative to delivery period) correction of inventory levels among freezers thus reducing wastage.

In the case of load balancing, the individual inventory levels, the demand and the delivery/usage levels for all freezers can be combined. In contrast, when individual inventory controllers are used the demand is fulfilled only when it is a positive value. To prevent excess inventory being shipped back, shifting the load between the freezers can offset the negative demand (excess storage).

When there is no negative demand, the total demand is equal for both scenarios. However, if there is at least one freezer with negative demand, the corresponding excess inventory will be used to fulfil the positive demand of the other freezers at that echelon thus reducing the total order or target demand. Thus, the total order in the case of load balancing scenario will be smaller or at least equal to the total demand observed in the case of no load balancing.

In other words, the total error in inventory levels for load balancing scenario is smaller or equal to the scenario when the total error in inventory levels without load balancing.

48.2.2 Antenna Configuration

The selection process for finding the necessary number of antennas and their localization inside the freezer in order to achieve full visibility is illustrated in Fig. 48.3. First, a realistic RF coverage pattern has to be obtained for the utilized antennas in the target environment. Next, the individual antenna coverage pattern is used to determine the number and positions of antennas that provide the RF coverage of the entire freezer cavity. The selection of location and the number of antennas required for attaining such high read rates with low reading times is performed iteratively by adding and positioning one antenna at a time. When the final antenna configuration is acquired, hardware tests can be used to verify that

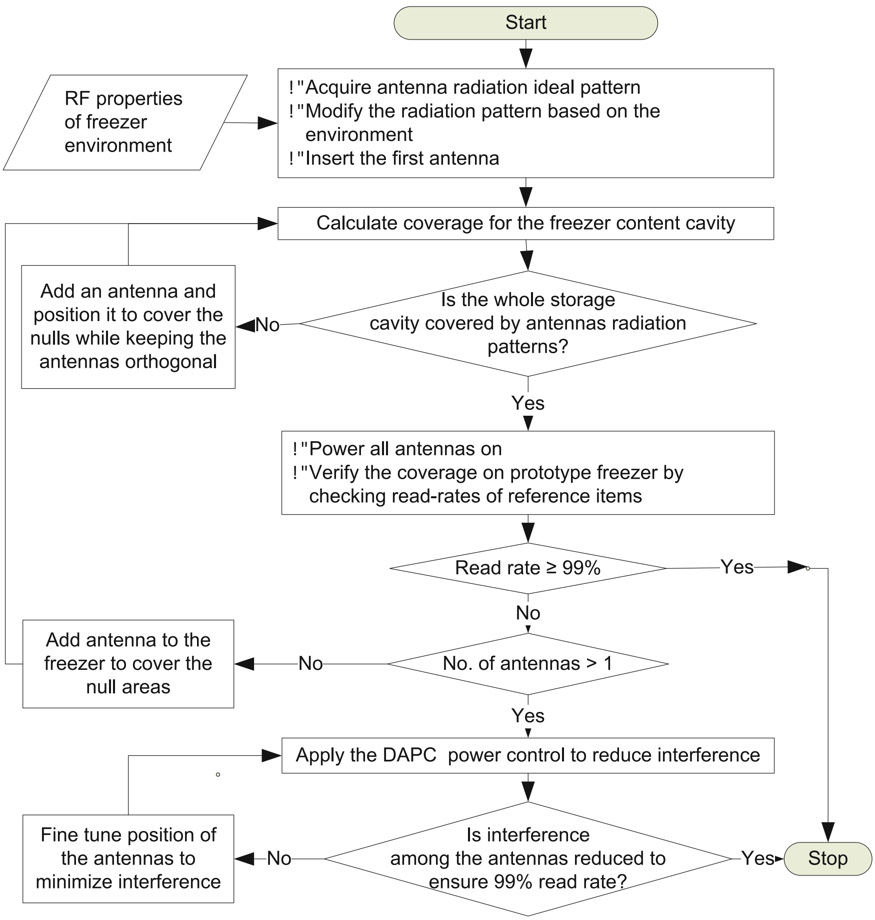


Fig. 48.3 Antenna selection configuration for Missouri S&T smart freezer

each antenna reads all the tags within its designated area. Finally, the proposed power control scheme is applied online in order do minimize interference among the antennas while ensuring the desired coverage inside the freezer. The proposed scheme to obtain antenna configuration is described next.

- **Acquisition of antenna radiation pattern**—The antenna radiation patterns [15] and propagation models are utilized to determine the antenna coverage.
- **Selection of antenna configuration to cover the freezer**—Starting with a single antenna, we calculate the coverage for the entire freezer. The freezer volume is divided into a grid of reference points spaced at predefined intervals and the signal strength is calculated from each RFID antenna. The employed propagation model has to include consideration of the absorption properties of the stored materials. More antennas are added to cover nulls (where the tag

cannot be read) when they occur. Each antenna is placed to provide best coverage based on their radiation pattern. Our experimental studies showed that the orthogonal placement of antennas minimize the final number of required antennas. When a new antenna is added, only the location of this antenna is varied, since the location of the existing antennas had been already established.

- **Testbed validation of the antenna configuration**—The overall performance of the freezer setup is validated on the hardware. The reference tags are placed at worst possible locations (e.g. in the middle of the item stack, at the corner of the designated cavity). The antennas are powered one at a time and the reference tags are read.
- **Fine-tuning of antennas location to reduce interference**—Due to limited space inside a freezer, the coverage obtained from the modeling steps may not be accurate due to the interference/collisions from other antennas operating simultaneously. We minimize the interference by using a locally asymptotically stable distributed power control (LASDPC) scheme or simply DPC. The DPC selects suitable power while ensuring there is no overlap in RF coverage among the antennas at all times. This essentially eliminates the collisions among the antennas. The proposed scheme ensures both bounded [14] and asymptotic stability. If the LASDPC alone cannot mitigate the interference, then the antenna position is fine-tuned.
- **Final configuration is achieved**—Once the consistent 99 % read-rate is achieved, the current configuration of antennas becomes the solution.

48.3 Simulation and Experimental Results

48.3.1 Simulation Study

In order to verify the effectiveness of the proposed inventory control scheme, several simulation scenarios were completed in Matlab. The simulation scenarios depict an industrial shop floor with 8 freezers. Actual demand for each freezer was modeled using an assortment of functions, for example a step function and a sinusoidal function. Demand variation $e_{dij}(k)$ was implemented using a uniform random variable with range $(-10, 10)$. Total error in the inventory level and total delivery request were selected as the performance metrics. The overall goal is to reduce the total error since high error leads to wastage of materials due to overstocking, or disruption of production due to insufficient inventory. Additionally, the total delivery request is observed in order to identify when excess amount of items is ordered.

The proposed scheme is compared with exponential weighted moving average (EWMA) method [2] with smoothing parameters $\alpha = (0.9)$ and $\beta = (0.1, 0.9)$. Total of 3 cases were simulated, with and without the distributed inventory control scheme and inventory balancing. Each scenario was repeated 500 times and the

Table 48.1 Simulation results

	Proposed scheme		EWMA $\alpha = 0.9$, $\beta = 0.1$		EWMA $\alpha = 0.9$, $\beta = 0.9$	
	Total error in inv. level	Delivery request	Total error in inv. level	Delivery request	Total error in inv. level	Delivery request
No—DIC	15,029	25,107	11,476	24,696	12,282	24,711
DIC w/o inventory balancing	9,833	24,674	9,860	24,723	10,347	24,722
DIC w/o inventory balancing	2,989	24,718	2,982	24,702	3,880	25,867

results were averaged in order to account for the effects of the stochasticity in the system.

Table 48.1 summarizes the results of the simulation study. The proposed scheme performs on par with the best-case scenario using EWMA. However, the EWMA parameters need to be manually selected based on experiments [2] since they have a direct effect on the accuracy of the demand prediction. In contrast, the proposed demand prediction model tunes these parameters online and without user intervention in order to reduce the prediction error.

The DIC backpressure scheme reduces the total error in inventory level by up to 15 % for EWMA and 34 % for the proposed scheme over the No–DIC scenarios since the ordering policy for the supply chain compensates for demand fluctuations. Further performance improvement is observed when the inventory load balancing method is introduced, for instance the errors in inventory levels are reduced by 62–70 % when inventory load balancing is added to the DIC scheme.

48.3.2 Hardware Experiments

In the hardware portion of the experimentation, impact of RFID reader and antenna types, number and placement of antennas, reading times and operational temperature on the read rate were investigated. Each experimental scenario was repeated at least five times to ensure consistency in the results. The containers were kept in the freezer for a prolonged time to study impact of exposure to low temperatures and ice buildup.

Placement of antennas inside the freezer that can deliver 99 % read rates were identified using the proposed methodology in Section II.C. The final configuration considered ferrous material storage in low temperature environment with frosting. The fine-tuned locations of 6 antennas are shown in Figs. 48.4 and 48.5.

The reader was operated with a maximum transmission power of 30 dBm. Tracked items are tagged with AD-220 Gen 2 tags from Avery Dennison were attached to the containers. The proposed LASDPC scheme was implemented at the

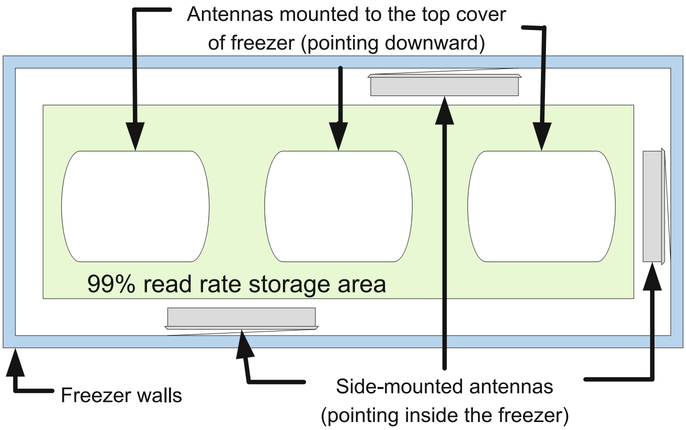


Fig. 48.4 RFID antenna locations for smart freezer solution

Fig. 48.5 RFID antenna location for smart freezer solution

reader. The antenna configuration design and confined freezer area resulted in the power control scheme operating in a simplified on/off mode, where the power control combats the multipath fading alone.

The experimental setup used in this work comprised of: AR-400 Matrics Class 1 Gen 2 RFID reader with dual directional antennas, and ALR-9800 Alien Class 1 Gen2 RFID reader with circularly polarized antennas. The freezer was loaded with 276 containers including 120 containers with non-RF-friendly (iron oxide) materials. Non-RF friendly materials absorb the RF signal and negatively affect the overall read rates.

Series of experiments have been conducted in order to evaluate the impact of each step of the proposed antenna configuration design scheme. First, a baseline scenario with simple antenna placement is presented. Next, the steps to determine the desired antenna number and placement are applied taking into account signal attenuation due to ferrous materials, frost, and tag orientation. Finally, the proposed power control is applied and tested in varying operating conditions. The worst case of −100 °F operating temperature has been thoroughly tested over a period of 3 months.

48.3.2.1 Experiments with Various Antenna Configurations

Table 48.2 was obtained during the initial testing of RFID-enabled freezer. The study shows inadequate read rates for simple antenna configuration for both Alien and MATRICS antennas. In general, the read rates increase with the number of antennas since the additional antennas can be positioned to cover the nulls. However, read times appear to be quite high which can be mitigated using DPC as presented in Table 48.3.

Moreover, the read rates improved for the 'multi-static' Alien antennas since they can switch between transmission and reception whereas a Matrics antenna includes two static elements. As a result, each Alien antenna act as two sets of fixed function antennas although physically there is one. This information should not be confused in the Table 48.2 between the manufacturers. Next, the proposed design methodology from Section II was applied and experimentally tested.

By using the flow chart in Fig. 48.3, the adequate number of antennas and their location was calculated to be equal to six and their respective locations are shown in Figs. 48.4 and 48.5.

Table 48.2 Comparison of matrics and Alien equipment

Reader	Antenna	Duration (s)	Read rate		
			1-Antenna (%)	2-Antennas (%)	4-Antennas (%)
Matrics AR-400	AN-100 dual directional	180	34–50	68–75	82–86
Alien ALR-9800	ALR-9610 circular	180	59–63	75–77	84–88

Each Matrics antenna include a separate transmitting and receiving element

Table 48.3 Read rates

Reader and antennas	# of antennas	Duration (s)	Temp. (°F)	No power control random placement (%)	Power control random placement (%)	Power control & antenna placement (%)
Alien Alr-9800 ALR-9610 circular	6	40	20	81	94	99
			−20	80	93	99
			−40	78	93	99
			−60	75	89	99
			−100	73	88	99

48.3.2.2 Results for Varying Temperature

The selected antenna configuration was validated at several operating freezer temperatures with proposed antenna placement and power control. The temperature varied from 20 to −100 °F, which is a typical range of temperatures used in the composite industry. Table 48.3 presents the summary of results. Missouri S&T scheme renders 99 % read rates regardless of freezer temperature since the methodology ensures coverage in the presence of severe fading and interference while ensuring a low read time of 40 s.

48.4 Conclusions

This paper introduces a novel DIC scheme and a methodology to identify the location and number of antennas required for attaining full visibility. The proposed DIC scheme with inventory balancing reduces errors in inventory level when contrasted against a simple demand prediction scheme. The performance of the DIC scheme was demonstrated through Lyapunov approach. In order to achieve the 99 % read rates, Missouri S&T smart freezer solution was utilized. The experimental results demonstrate 99 % read rates within 40 s at −100 °F temperature with six antennas. Overall, the DIC scheme in concert with the smart freezer design delivers a reliable and efficient supply chain solution.

References

1. Jun HB, Kiritsis D, Xirouchakis P (2007) Research issues on closed-loop PLM. Comput Ind 58(8–9):855–868
2. Mills-Harris MD, Soylemezoglu A, Saygin C (2007) Adaptive inventory management using RFID data. Int J Adv Manuf Technol 32(9):1045–1051
3. Chow HKH, Choy KL, Lee WB, Lau KC (2006) Design of a RFID case-based resource management system for warehouse operations. Expert Syst Appl 30(4):561–576
4. Lian X, Zhang X, Weng Y, Duan Z (2007) Warehouse logistics control S management system based on RFID. In: Proceedings of IEEE international conference on automation and logistics, pp 2907–2912
5. Budak E, Catay B, Tekin I, Yenigun H, Abbak M, Drannikov S (2007) Design of an RFID-based manufacturing monitoring and analysis system. In: Proceedings of the 1st annual RFID Eurasia, pp 1–6
6. Huang GQ, Zhang YF, Jiang PY (2007) RFID-based wireless manufacturing for walking-worker assembly islands with fixed-position layouts. Robot Comput-Integr Manuf 23(4):469–477
7. Imburgia MJ (2006) The role of RFID within EDI: building a competitive advantage in the supply chain. In: Proceedings of IEEE international conference on in service operations and logistics, and informatics, pp 1047–1052

8. Hinske S (2007) Determining the position and orientation of multi-tagged objects using rfid technology. In: Proceedings of the 5th IEEE international conferece on pervasive computing and communications workshops, pp 377–381
9. Yang X, Moore PR, Wong CB, Pu JS (2007) A Component-based software framework for product lifecycle information management for consumer products. IEEE Trans Consum Electron 53(3):1195–1203
11. Decker C, Kubach U, Beigl M (2003) Revealing the retail black box by interaction sensing. In: Proceedings of the 23rd international conference on distributed computing systems, p 328
12. Blau J (2004) Supermarket's futuristic outlet. IEEE Spectr 41(4):21–25
13. Bornhovd C, Lin T, Haller S, Schaper J (2004) Integrating automatic data acquisition with business processes experiences with SAP's auto-ID infrastructure. In: Proceedings of the thirtieth international conference on very large data bases, vol 30.p 1182–1188
14. Clarke RH, Twede D, Tazelaar JR, Boyer KK (2006) Radio frequency identification (RFID) performance: the effect of tag orientation and package contents. Packag Technol Sci 19(1):45–54
15. Soylemezoglu A, Zawodniok MJ, Cha K, Hall D, Birt J, Saygin C, Sarangapani J (2006) A testbed architecture for Auto-ID technologies. Assembly Autom 26(2):127–136

Chapter 49
Advances Toward Sustainability in Manufacturing

E. Gilabert, A. Arnaiz and O. Revilla

Abstract This paper addresses the need of developing sustainable strategies by taking advantage of e-platforms bringing advanced asset management solutions. Even though most asset management systems are cost oriented, it is also clear that new factors are gaining weight when considering strategies and operation, like safety or environmental issues. In this sense, LCA methodology enables to assist an effective integration of the environmental considerations in the decision-making process, and this paper presents different applications developed in this address. The paper also shows an e-platform based on current maintenance technologies already developed, together with strategies to select the most cost-efficient path in the search for a maintenance 'excellence' strategy. ICT technologies impact both the economical and environmental cost of the manufacturing processes or the factory as a whole. An integrated strategic decision tool for sustainability must be able to simulate and assess cost variations in processes and business models and also incorporate novel metrics, business concepts, and paradigms in order to estimate adequately the interactions between factory, production systems, and processes.

49.1 Introduction

For most manufacturing companies the concept of sustainability is impacting their processes, making it a considerable business challenge, especially concerning the environmental dimension, as nearly all manufacturing models are currently based

E. Gilabert (✉) · A. Arnaiz · O. Revilla
Fundación Tekniker, Avda Otaola 20, Eibar 20600, Spain
e-mail: egilabert@tekniker.es

A. Arnaiz
e-mail: aarnaiz@tekniker.es

O. Revilla
e-mail: orevilla@tekniker.es

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_49,

on the old paradigm of unlimited resources and unlimited capacity for regeneration. The manufacturing Industry within Europe is driven by economic incentives, particularly in the context of the recent world recession, to implement significant energy efficient improvements in its processes and the equipment it uses (electric motors, compressors, etc.). European and national legislation impacts on the energy consumption and in that connection industry must also take the necessary measures to fulfill the gap on energy consumption limitations imposed by European and National legislation and on greenhouse gas emissions imposed by the National Allocation Plans as foreseen in the Emissions Trading Directive. In addition, companies are under pressure to respond to rising energy costs and the need to protect the environment. Growth in energy consumption has a direct impact on the deterioration of the environment and on climate change. Air quality is a major environmental concern for the EU. The Commission is currently elaborating the EU Clean Air Programme (CAFE), where the harmful effects of ozone and especially particulates are revealed for human health and ecosystems.

Given this, it is important to focus on the need of developing sustainable and energy efficient manufacturing strategies through the research, develop a demonstration of novel methodologies and technologies allowing the improvement of environmental, economic, and social policies. In this sense, asset management is crucial to manufacturing operations. In many firms the facilities and the production equipment represent the majority of invested capital, and deterioration of these facilities and equipment increases production costs, reduces product quality, and has a significant impact on energy consumption. Over recent years, the importance of asset management, and therefore maintenance management within European manufacturing organizations has grown [1, 2].

On the other hand, there exist a number of technologies, processes and even business models that may bring great benefits to a manufacturing company in its search for greater eco-friendliness and competitiveness. There exist many different sources to break down different Best Available Techniques in every machine, process, and industry. In many cases, it is difficult to demonstrate the feasibility of the return on investment related to the new systems and to give priority between selected improvements, which stops their implementation. But there is a direct way to address this challenge by developing a strategic decision tool that simulates the ecological footprint of the factory as a whole being able to perform a bottom approach from factory and production systems to process.

49.2 Life Cycle Assessment in Manufacturing

Over the past decade, asset management has evolved to be one of the most important areas within manufacturing organizations, large or small. The success and sustainability of any energy intensive industry is measured by their ability to perform well with regard to certain criteria such as cost, quality, delivery, dependability, innovation, and flexibility. In order to achieve and maintain these criterions these organizations are undertaking efforts to improve quality and productivity and reduce manufacturing costs by examining the activities of the asset management function. It is critical because the goal is to extend equipment life, improve equipment availability, and retain equipment in a clean, reliable, safe, and energy efficient condition. Customers demand ever-increasing reliability, with faster lead and delivery times.

Existing strategies seek for optimized operational solutions involving asset management costs, energy efficiency, and CO_2 emissions. On the other hand, the European Union set itself ambitious targets by the year 2020 (to reduce the output of greenhouse gases by 20 %, to improve energy efficiency by 20 %, and to increase the percentage of renewable energy by 20 %).

Even though many tools for the assessment of environmental impacts are available, such as Life Cycle Assessment, Material Flow Analysis (MFA), and Environmental Impact Assessment (EIA), Life Cycle Assessment is the most widely used technique that allows organizations to identify the strong correlation between energy efficient maintenance and productivity. Using LCA when making purchasing and maintenance decisions ensures that all benefits are taken into account.

In adequate conditions, the LCA methodology enables to assist an effective integration of the environmental considerations in the decision-making process. The Life Cycle Assessment (LCA) and Life Cycle Cost (LCC) concepts are directly related to the sustainability of the manufacturing processes [3]. The objective of use LCA and LCC is twofold. On one hand, LCA model is able to analyze the complex interaction between the product and manufacturing process and the environment, considering the energy efficiency, from cradle to grave, while the LCC can analyze the total "lifetime" cost to purchase, install, operate, maintain and dispose of the products.

Using LCA would result in a more realistic, integrated, and accurate view of the potential optimization of a production line. In addition it would be very useful in assisting the whole product development by identifying more sustainable options in process selection, design, and optimization, and in particular the energy efficiency and CO2 emissions. Therefore, LCA as analytical tool allows the assessment and simulation of the environmental impacts that different operational decisions, as well as monitoring technologies [4], have on the whole life cycle of the manufacturing plant.

New regulations are pushing companies to introduce more sophisticated management tools to drive the environmental change. Different tools have been developed in this address, although more effort must be carried out.

49.2.1 Ecodesign in Energy Intensive Sector

The cement sector is making an effort in protecting the environment, optimizing the usage of facilities, equipment upgrading, investment in means of reducing environmental impacts, and offering the possibilities of productive processes for waste recovery generated in other activities. In this sense, the cement industry works in the direction of sustainable development. In this context, the Basque funded project ECOCEM has developed a decision tool capable of simultaneously applying strategies to waste management and strategies for sustainable management resources in the cement industry in order to minimize and optimize the environmental impact.

The ECOCEM application offers the possibility to assess the economic and environmental impacts generated during the most critical manufacturing step in the manufacturing of cement: the rotary kiln. Thus, users can obtain the environmental impacts and costs associated to the process, as well as simulate and analyze the environmental impacts and costs that would be generated by modifying the percentages of different raw materials or fuels, or introducing a new treatment system. The project ECOCEM sets a methodology combining simulation and analysis of environmental impacts (based on the methodology providing information on best techniques or practices that would reduce environmental impacts critics.

49.2.2 Life Cycle for Building Machine Tools

The EU-founded collective research project PROLIMA (Environmental Product Life Cycle Management for Building Competitive Machine Tools) was developed focusing on the sustainability of the European producers of machine tools, which try to increase the competitiveness based on quality, value for money, and low environmental impact. PROLIMA Sustainable Machine Tool Design Methodology (SMTDM) aims to provide a global methodology for sustainable design, manufacturing, operation, and maintenance considering the economical, environmental, and social issues (Fig. 49.1).

PROLIMA's core is their Decision Support System (DSS) which integrates the information of the Life Cycle Cost (LCC) tool and the Life Cycle Assessment (LCA) tool, together with Social aspects (safety, ergonomics, aesthetics, usability,…) and functional aspects (productivity, precision, cost per piece, etc.). Requirements models can be customized to each kind of product according the characteristics of the manufacturer and the customer. As an example, Machine Tool Sustainability Index (MTSI) was developed for machine tool sector. According to these models, designers can evaluate the Sustainability Index of different configurations from early design stages, when most of critical decisions are made and main costs are committed. Using an appropriate methodology supported by useful tools give the pragmatic guidance to choose the best ecodesign considering manifold points of view over the product life span.

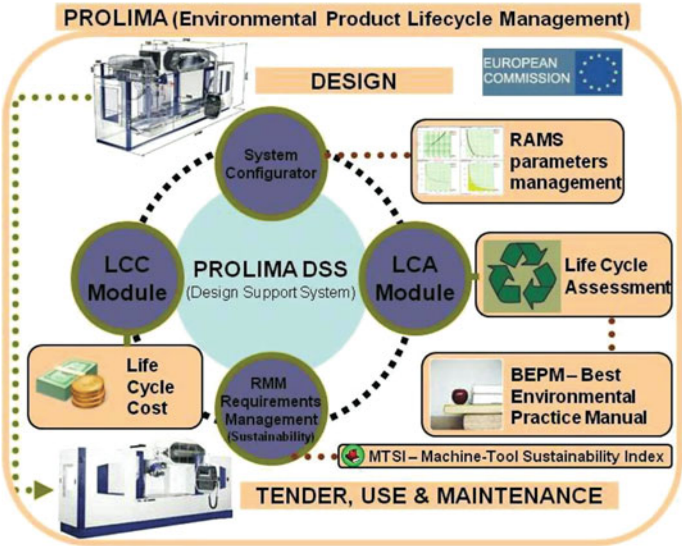


Fig. 49.1 PROLIMA concept

49.2.3 Ecodesign Tool for EEE

Electrical and Electronic Equipment (EEE) is developing fast and spreading over every part of modern life. In the European Union, the current Waste Electrical and Electronic Equipment (WEEE) is the fastest growing main stream, which is three times faster than average waste. WEEE has been identified as a priority area to take specific measures on a European scale Because of the risk to human health and effect on environmental change. The main purpose of WEEE-NET project is the demonstration of the feasibility of an innovative platform based on ICT, for the sustainable management, reuse, and recycling of WEEE in order to promote and support the implementation of Community Environmental policy on e-waste. The project targets on electrical appliances sector and includes all involver stakeholders, from manufacturing industry to the consumer. The project also focuses on ecodesign of electrical and electronic equipment, taking into account the whole Life Cycle, from manufacturing to waste treatment. The tool enables the monitoring of the product in order to facilitate the proper management when the useful life is close to end.

49.3 E-Platform for Sustainability

The integration of novel joint LCA/LCC/RAMS analysis tools can be used for simulation and evaluation purposes for the new advanced maintenance techniques and technologies. The recently finished EU-funded Integrated Project DYNAMITE (Dynamic Decisions in Maintenance) has brought together a group of technologies that can be integrated in a structured way, yet flexible enough to allow the selection of a particular subset of the technologies. The main idea behind the platform is a series of "plug and play" hardware and software components that can be included on demand in any existing use case, without a need to remove any legacy system. The growth of wireless communication technology, mobile technology, and web technology have allowed maintenance strategies to be developed using accurate information collect using sophisticated technologies [5, 6]. Technological developments in e-maintenance systems, radio-frequency identification (RFID), and personal digital assistant (PDA) have proven to satisfy the increasing demand for improved machinery reliability, efficiency, and safety [7]. Maintenance task selection is now developed by applying a blend of leading-edge communications and sensor technology including Radio Frequency Identification (RFID) and Personal Digital Assistant (PDA) to enhance diagnostic and prognostic capabilities [8–10].

In this sense, it is important to insist on the distributed nature of the platform, which facilitates the inclusion of new modules (i.e. metering devices for energy-efficiency indicators, novel and existing algorithms for EMS decision-support). This maintenance operating platform should be also combined innovative energy data Smart Metering systems have been developed. This innovative approach to energy metering allows for two-way communication between the utility (supplier or Distribution System Operator [DSO]) and the meter. Smart meters are modern, innovative electronic devices capable of a wide range of useful information, enabling the introduction of new energy services.

This e-platform helps industry to monitor, manage, and optimize their energy usage for maximum efficiency and cost savings. In addition the software would include reporting and analysis tools that evaluate the energy use patterns of all processes and pinpoint areas for improvement. Finally it would be able to coordinate electricity costs to provide a true cost. And with LCA/LCC analysis tools we will evaluate the best strategy for the sustainable and cost-effective manufacturing plant.

49.4 Conclusions

The need of developing sustainable strategies by taking advantage of e-platforms and bringing advanced asset management solutions has been addressed. New factors such as safety or environment issues are gaining weight in asset management. LCA methodology enables to assist an effective integration of the environmental considerations in the decision-making process, and this paper presents different applications developed in this approach. The paper also shows an e-platform based on current maintenance technologies already developed, together with strategies to select the most cost-efficient path in the search for a maintenance 'excellence' strategy. ICT technologies impact both the economical and environmental cost of the manufacturing processes or the factory as a whole. An integrated strategic decision tool for sustainability must be able to simulate and assess cost variations in processes and business models and also incorporate novel metrics, business concepts, and paradigms in order to estimate adequately the interactions between factory, production systems and processes.

References

1. Arnaiz A, Maurtua I, Susperregi L, Arana R (2004) Maintenance: future technologies. In: Proceedings of the intelligent manufacturing system (IMS) international forum. Como, Italy, pp 300–307
2. Crespo Marquez A, Gupta JND (2006) Contemporary maintenance management: process, framework and supporting pillars. Omega 34(3):325–338
3. Takata S, Kimura F, van Houten FJAM, Westkämper E, Shpitalni M, Ceglarek D, Lee J (2004) Maintenance: changing role in life cycle management. CIRP Ann: Manuf Technol 53(2):643–656
4. Jardine A, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech Syst Sig Process 20(7):1483–1510
5. Lee J (2001) A framework for web-enabled e-maintenance systems. In: Proceedings of the second international symposium on environmentally conscious design and inverse manufacturing. Milwaukee, WI, pp 450–459
6. Lee J, Ni J, Djurdjanovic D, Qiu H, Liao H (2006) Intelligent prognostics tools and e-maintenance. Comput Ind: Spec Issue e-Maintenance 57(6):476–489
7. Muller A, Crespo A, Iung B (2008) On the concept of e-maintenance: review and current research. J Relib Eng Syst Saf 93:1165–1187
8. Arnaiz A, Emmanoulidis C, Iung B, Jantunen E (2006) Mobile maintenance management. J Int Technol Inf Manag 15(4):11–22
9. Arnaiz A, Gilabert E, Jantunen E, Adgar A (2009) Ubiquitous computing for dynamic condition based maintenance. J Qual Maintenance Eng (JQME). 15(2):151–166
10. Legner C, Thiesse F (2006) RFID based maintenance in Frankfurt airport. IEEE Pervasive Comput 5(1):34–39

Chapter 50
Wavelet Analysis and Fault Feature Extraction of Rolling Bearing

Z. Yang and L. Gao

Abstract Aiming at inner ring single fault and inner ring, outer ring compound fault of rolling bearing, wavelet analysis and nonlinear redundant lifting wavelet packet analysis are introduced in this paper to process the fault vibration signals in order to realize fault diagnosis of bearing. Conclusions can be drawn from the results that spectrum analysis is of limitation for fault feature extraction, while wavelet analysis and nonlinear redundant lifting wavelet packet analysis are more effective in extracting fault feature information and the latter has more advantage in compound fault analysis.

50.1 Introduction

Rolling bearing is one of the widely used components of large equipment, and it is also classified as a critical component because of its high frequency of failure. Once fault occurs, normal operation of equipment will be effected which can lead to huge economic loss and even personnel injury. Therefore, condition monitoring for rolling bearing is of great practical significance.

Many signal analysis and processing methods are further investigated by researchers so as to realize effective condition monitoring of rolling bearing. Among them, wavelet analysis based on multi resolution analysis has been widely applied [1]. According to the deficiency that classical wavelet construction depends on Fourier transform and the shape of wavelet is fixed, the lifting

Z. Yang (✉) · L. Gao
Key Laboratory of Advanced Manufacturing Technology, Beijing University of Technology, Chao Yang District, Beijing 100124, China
e-mail: zijing396@sina.com

L. Gao
e-mail: lead0003@163.com

J. Lee et al. (eds.), *Engineering Asset Management 2011*, 571
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_50,

algorithm was proposed by Sweldens [2] in 1995 to construct wavelet with expected characteristic in time domain to match signal feature. Subsequently, nonlinear thought was presented by Claypoole et al. [3] based on the characteristics of lifting wavelet, that is, to select different predict operators on the basis of local characteristics of images.

50.2 Wavelet Analysis

Wavelet has excellent localization characterization capacity both in time domain and frequency domain. With the binary discrete wavelet transform based on Mallat fast algorithm, signal is decomposed layer by layer into low frequency approximation signal and high frequency detail signal by conducting inner product operation with two-channel filter banks. Then next, the transient components in signal can be effectively extracted by further de-noising and reconstruction processing. Concrete steps of wavelet analysis adopted in this paper are as follows:

1. The frequency band where high frequency peaks locate is determined by spectrum analysis.
2. Multilayer wavelet decomposition is performed to get scale coefficients and wavelet coefficients under different scale.
3. Wavelet coefficients corresponding to frequency which are most similar to the frequency band where high frequency peaks locate are selected to conduct single branch reconstruction.
4. Hilbert demodulation of reconstructed signal is performed.
5. Fault recognition is conducted according to the analysis of frequency in demon spectra.

50.3 Nonlinear Redundant Lifting Wavelet Packet Analysis

The decomposition process of redundant lifting wavelet transform contains two steps: 1) predict and 2) update, through which high frequency detail signal and low frequency approximation signal are respectively obtained:

$$d(k) = x(k) - P[x(k)], \tag{50.1}$$

$$a(k) = x(k) + U[d(k)], \tag{50.2}$$

where $x(k)(k \in Z)$ is original signal, $P[\cdot]$ is predict operator and $U[\cdot]$ is update operator, while $d(k)$ and $a(k)$ are detail signal and approximate signal obtained respectively through prediction and updating.

The reconstruction process has three steps: undo update, undo predict and merge:

$$x(k) = \frac{1}{2} \cdot [x^u(k) + x^p(k)], \tag{50.3}$$

where $x^u(k)$ and $x^p(k)$ are the samples respectively received by prediction recovery and updating recovery.

In this paper, nonlinear thought is introduced. Different predict operators and update operators are chosen for all the nodes of wavelet packet to realize decomposition and reconstruction. Let N and $\tilde{N}$are respectively the length of predict operator and update operators. When N was small, the frequency characteristics of the scale function and wavelet function cannot be improved even if $\tilde{N}$ is increased; while when N increased gradually, the frequency characteristics of the scale function and wavelet function will be improved. When $N$is not increased enough, the frequency characteristics of the scale function and wavelet function will not improve much [4]. Therefore, six group different wavelet functions $\left(N, \tilde{N}\right)$ (4, 4), (12, 4), (20, 4), (12, 12), (20, 12), and (20, 20) are chosen to perform the wavelet packet decomposition of signal. Meanwhile, l^p norm is adopted as the evaluating parameter of sparsity of decomposed coefficients. The l^p norm is defined as follows:

$$\|x\|_p = \left(\sum_k |x_k|^p\right)^{1/p}, \quad p \le 1, \tag{50.4}$$

The more similar the interesting components in signals are to the wavelet function, its corresponding wavelet coefficients will be larger, while the wavelet coefficients for other components in signals will be smaller or even approach zero, which means the coefficients become sparser and the corresponding l^p norm is lower. Hence, the predict operator and update operator corresponding to wavelet coefficients with the lowest l^p norm are chosen as the optimal one. Then, node-signal single branch reconstruction is completed to finally realize the nonlinear redundant lifting wavelet packet analysis [5]. Concrete steps are as follows:

1. Six group different wavelet functions are chosen to perform redundant lifting wavelet packet decomposition of signal.
2. Normalized l^p of coefficients acquired by decomposition are computed to determine the optimal predict operator and update operator.
3. Wavelet packet energy analysis of all nodes obtained by decomposition with optimal operators is conducted.
4. Single branch reconstruction of node with maximum energy is performed followed by Hilbert decomposition.
5. Fault recognition is conducted according to the analysis of frequency in demon spectra.

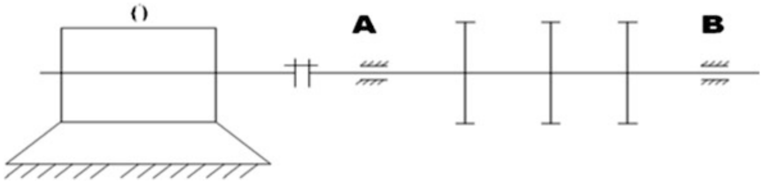


Fig. 50.1 Diagram of test-bed of bearing

50.4 Experimental Data Analysis

Vibration signals of both inner ring single fault and inner ring, outer ring compound fault of bearing are collected from test-bed of bearing whose diagram can been seen in Fig. 50.1: shaft is driven by motor through coupling to bring along the rotation of three rotors. A and B are rolling bearings of type 6,307, while bearings with different fault can be replaced at the end of B. Motor speed is set at 1,496 r/min according to which fault characteristic frequency of outer ring and inner ring can be respectively calculated as 76.08 and 123.386 Hz.

During the signal acquisition, sampling points is set as 8,192 while sampling frequency is 15,360 Hz. Analytical results are as follows:

50.4.1 Inner Ring Fault of Bearing

The three subgraphs in Fig. 50.2 are respectively the time domain chart, spectrogram, and wavelet packet energy analysis chart of signal with inner ring fault.

Firstly, periodic impact signal that indicates bearing fault can be seen in the time domain chart.

Secondly, wavelet analysis is performed. Wavelet coif4 with ideal compact support is selected to signal for triple-layer decomposition. As it can be seen from spectrogram that the main high frequency peaks are located at the frequency band of 3,358–4,076 Hz, so detail signal on the second layer (d2) is chosen for single branch reconstruction and demodulation.

Thirdly, nonlinear redundant lifting wavelet packet analysis is performed. As the wavelet packet energy analysis chart shows, node 5 with the maximum is chosen for node-signal single branch reconstruction and demodulation.

The three subgraphs in Fig. 50.3 are respectively the local spectrogram, wavelet analysis and its demodulation spectra, nonlinear redundant lifting wavelet packet analysis and its demodulation spectra of signal with inner ring fault.

125.6 Hz which is close to inner ring fault characteristic frequency 123.386 Hz can be found in the local spectrogram.

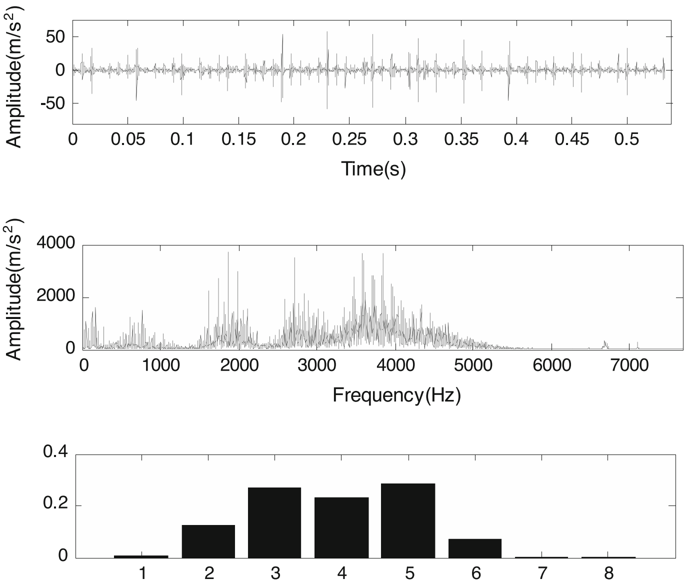


Fig. 50.2 Signal with *inner ring* fault: time domain chart (*top*), spectrogram, and wavelet packet energy analysis chart (*bottom*)

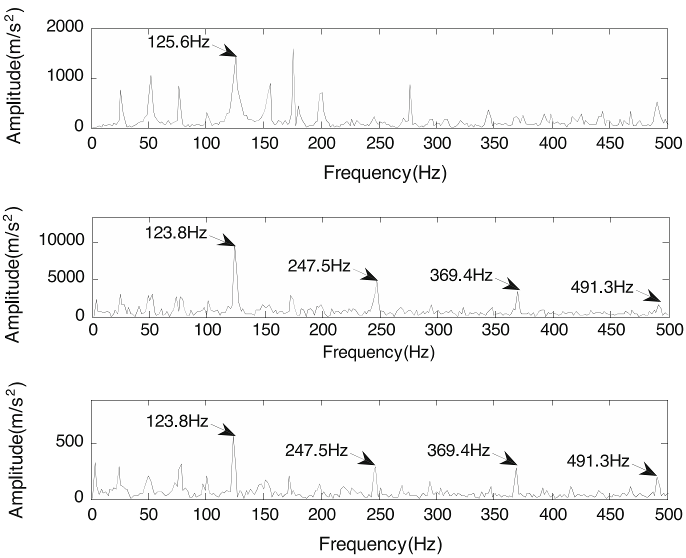


Fig. 50.3 Signal with *inner ring* fault: local spectrogram (*top*), wavelet analysis and its demodulation spectra, and nonlinear redundant lifting wavelet packet analysis and its demodulation spectra (*bottom*)

123.8 Hz and its multiple frequency 247.5, 369.4, and 491.3 Hz that in accordance with inner ring fault can be seen in the wavelet analysis and its demodulation spectra.

The same frequency 123.8, 247.5, 369.4, and 491.3 Hz can also be clearly seen in the nonlinear redundant lifting wavelet packet analysis and its demodulation spectra.

Therefore, one conclusion can be made that compared with spectrum analysis, wavelet analysis and nonlinear redundant lifting wavelet packet analysis not only have higher accuracy in extracting the basic frequency of inner ring fault characteristic frequency, but also can extract multiple frequency which leads to more effective fault recognition of rolling bearing.

50.4.2 Inner Ring and Outer Ring Compound Fault of Bearing

The three subgraphs in Fig. 50.4 are respectively the time domain chart, spectrogram, and wavelet packet analysis chart of signal with inner ring and outer ring compound fault.

Firstly, impact components can be seen in the time domain chart.

Secondly, wavelet analysis is performed. Wavelet coif4 is used to signal for triple-layer decomposition. As the main high frequency peaks are located at the

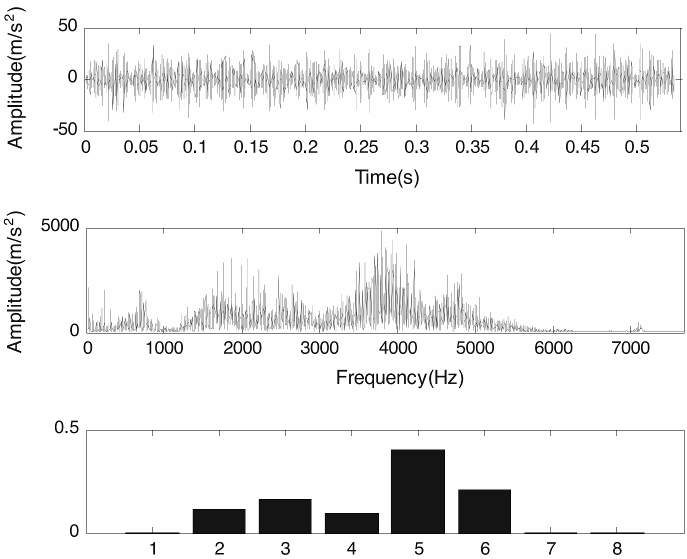


Fig. 50.4 Signal with *inner ring* and *outer ring* compound fault: time domain chart (*top*), spectrogram, and wavelet packet energy analysis chart (*bottom*)

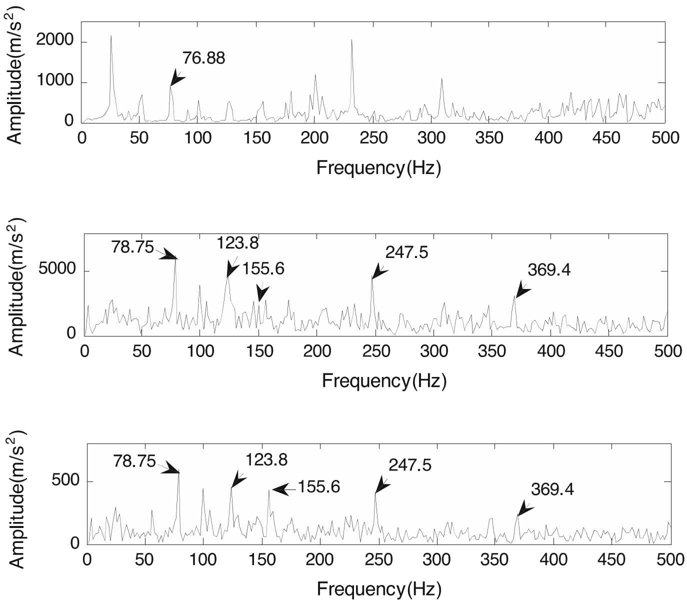


Fig. 50.5 Signal with *inner ring* and *outer ring* compound fault: local spectrogram (*top*), wavelet analysis and its demodulation spectra, and nonlinear redundant lifting wavelet packet analysis and its demodulation spectra (*bottom*)

frequency band of 3,579–4,299 Hz, so detail signal on the first layer (d1) is chosen for single branch reconstruction and demodulation.

Thirdly, nonlinear redundant lifting wavelet packet analysis is performed. Node 5 is chosen for node-signal single branch reconstruction and demodulation as it has the maximum energy.

The three subgraphs in Fig. 50.5 are respectively the local spectrogram, wavelet analysis and its demodulation spectra, nonlinear redundant lifting wavelet packet analysis and its demodulation spectra of signal with inner ring and outer ring compound fault.

Only 76.88 Hz which is close to outer ring characteristic frequency 76.081 Hz can be found in the local spectrogram.

Not only 123.8 Hz and its multiple frequency 247.5 and 369.4 Hz that in accordance with inner ring fault, but also 78.75 Hz and its multiple frequency 155.6 Hz that indicating outer ring fault can all be seen in the wavelet analysis and its demodulation spectra.

The same frequency 123.8, 247.5 and 369.4 Hz in accordance with inner ring fault, as well as 78.75 and 155.6 Hz indicating outer ring fault can also be clearly seen in the nonlinear redundant lifting wavelet packet analysis and its demodulation spectra.

In sum, wavelet analysis and nonlinear redundant lifting wavelet packet analysis have greater efficiency in compound fault feature extraction, while the double frequency of outer ring fault characteristic frequency 155.6 Hz extracted by the latter is much more clearly that of the former.

50.5 Conclusion

Wavelet analysis and nonlinear redundant lifting wavelet packet analysis are both introduced in this paper to process the vibration signal respectively with inner ring single fault and inner ring and outer ring compound fault. Results show that compared with spectrum analysis, wavelet analysis and nonlinear redundant lifting wavelet packet analysis are effective for fault feature extraction, while the latter is of more superiority in analyzing compound fault.

References

1. Mallat S (1989) A theory for multi-resolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell 11(7):674–693
2. Sweldens W (1996) The lifting scheme: a construction of second generation wavelets. SIAM J Math Anal 29(2):511–546
3. Claypoole RL, Geoffrey D, Sweldens W, Baraniuk RG (2003) Nonlinear wavelet transforms for image coding via lifting. IEEE Trans Image Proc 12(12):1449–1459
4. Duan, CD (2005) Research on fault diagnosis techniques using second generation wavelet transform. Xi'an Jiaotong University College of Mechanical Engineering, Xi'an Shanxi, p 510–513
5. Gao L, Yang Z, Cai L, Wang H, Chen P (2011) Roller bearing fault diagnosis based on nonlinear redundant lifting wavelet packet analysis. Sensors 11(1):260–277

Chapter 51
Failure Mode Analysis Based on MFM-HAZOP Model of Gathering System

J. Wu, L. Zhang, W. Liang and J. Hu

Abstract Gathering system is not only an important oilfield production facility system for gathering and transporting oil and gas but also is a key section to realize the function of oil and water separation. However, the high complex failure coupling relation among separate production process sections, personnel operation, and equipment leads to which is a complex high potential hazard and failure relational degree within subsystems production system, make it difficult to analyze failure mode. Therefore, it is urgent to demand a higher level effective operational failure analysis model to simplify the complex gathering system to resolve the realistic safety problem. A failure mode analysis of gathering system based on a novel MFM-HAZOP deep knowledge model is proposed. By decomposing goals, functions and components, a graphical MFM model of gathering system is established. Combined with HAZOP analysis technology, this paper sets up MFM-HAZOP model to perfect studying failure mode of gathering system and accurately analyzes reason of failures, which is the base of gathering system safety decision to guarantee the gathering system safely operating.

J. Wu (✉) · L. Zhang · W. Liang · J. Hu
College of Mechanical and Transportation Engineering, China University of Petroleum, Beijing 102249, China
e-mail: cup.wujing@gmail.com

L. Zhang
e-mail: zhanglb@cup.edu.cn

W. Liang
e-mail: lw@cup.edu.cn

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_51,

51.1 Introduction

Gathering system is a typical representative of a complex heterogeneous industrial system, which realizes function of collecting transportation separation purification and storage of gas and oil produced from wells. Due to the variability among equipment operation personnel and site environment, it is very important to research the effects of failures on other parts of the system when some components failures occur. Failure Mode Analysis, or FMA for short, is used to find all possible faults that may happen in a system and to discover effects or causality based on the correlations among faults. Traditionally, some basic methods have been studied for analyzing the cause-effect of failure modes. The classical procedure is failure mode and effect analysis (FMEA) to analyze potential failure modes within a system for classification by the severity and likelihood of the failures, see Refs. [1, 2]. To give a more logical reasoning between component faults and consequences, the fault tree analysis (FTA) is used to quantify more complex process failures, see Refs. [3, 4]. Another method is Goal Tree-Success Tree method (GTST) to model deep knowledge about complex industrial systems, particularly for the matter of fault diagnosis illustrating by the goal tree of system functions and the success tree of the physical structure and the relationships among variables, see Ref. [5]. However, these methodologies suffer from following problems:

1. It is difficult to detect several failures at the same time and find all possible ways a component may fail even though these models on the purpose of searching for all possibilities.
2. Because of manual analysis based on personal knowledge and experiences of experts, thus it is likely to miss important scenarios along with analysis.
3. If some processes being improved or equipment replaced in a system, it is urgent to ensure whether the existing identified possible faults are still valid. However, obviously, this question is difficult to be answered; it may require the whole analysis to be done once again.
4. Without reasoning rules, it is harder to reveal the coupling among failure modes and figure out which component faults can interact to be available for achieving system functions. Thereby, operators could fail to clarify the root causes of failure situation.

This article will focus on failure mode analysis using a graph theory model called multilevel flow models (MFM), which has many nice properties, such as possibility for on-line prediction of failures in real time, see Ref. [6] (Failure Mode Analysis Using Multilevel Flow Models), combining with HAZOP study to model MFM-HAZOP (Fig. 51.1) and give a failure mode analysis for gathering system discussed later. For more detailed discussion of related work and their relations to MFM-HAZOP, see Refs. [7, 8].

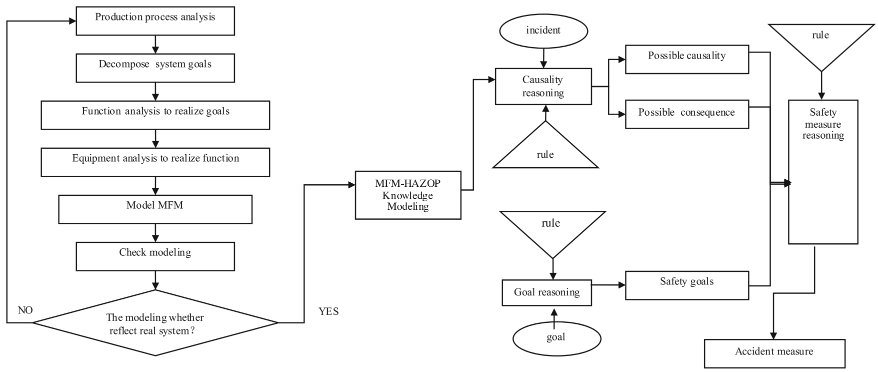


Fig. 51.1 How to build the MFM-HAZOP model

51.2 Multilevel Flow Models

51.2.1 Basic Theory of MFM

MFM [9] is a hierarchical structure modeling method based on target of a process. This approach attack problems from material flow, energy flow and information flow angles to abstract real physical system. Through using some specific graphic symbols, it describes the system goals, functions and components to model the production process of plant. In MFM, the goals are basis of modeling thought to realize functions of each part of system, such as "heat the oil" and the functions nodes (Fig. 51.2) consisting of function relate to goals to represent the capabilities of a system, for example, "transport oil", and the components represent the physical structures of a system, such as a piece of pipeline. Graph modeling (such as MFM), a deep knowledge, can express complex causality, which contains the ability of mass potential information, especially suitable for application in safety area. After invented by Lind [10] at the Technical University of Denmark, MFM has been proved to be effectively contributing for several diagnostic algorithms, such as measurement validation, alarm analysis, failure mode analysis, sensor fault detection, and fault diagnosis, see Refs. [6, 11–14].

In this article, the MFM modeling flow chart involves the following specific steps:

1. Comprehensively analyze the production process and fully know the prototype of system.
2. The goals are decomposed into a series of sub-goals.
3. Analyze and study the function characteristics of each part of system which aim at realizing the goals and list all conditions and limitations of function realization.
4. Find equipment components and its corresponding functional mapping relation.

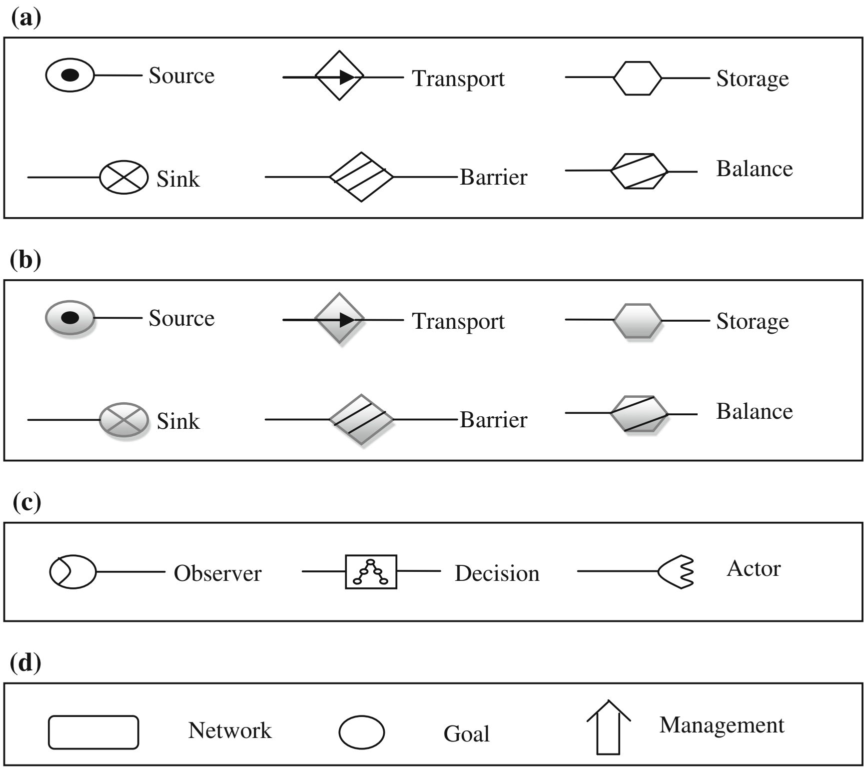


Fig. 51.2 MFM symbols for functions. **a** Mass flow function nodes. **b** Energy flow function nodes. **c** Information flow function nodes. **d** Network, goal, and management

5. Establish the multi-level flow models based on relation among goals, functions, and components.
6. Model inspection and correction.

51.2.2 Case Study

To illustrate how MFM modeling is used in gathering system, an example is presented in Fig. 51.3.

Gathering system is a main production facility in oilfield construction, which plays an important role in oil and gas production. Described in Fig. 51.3 there is a prototype of gathering system. This system process flow is that after being pre-hydrated and heated, the gas-oil mixture liquid enters into three-phase separator.

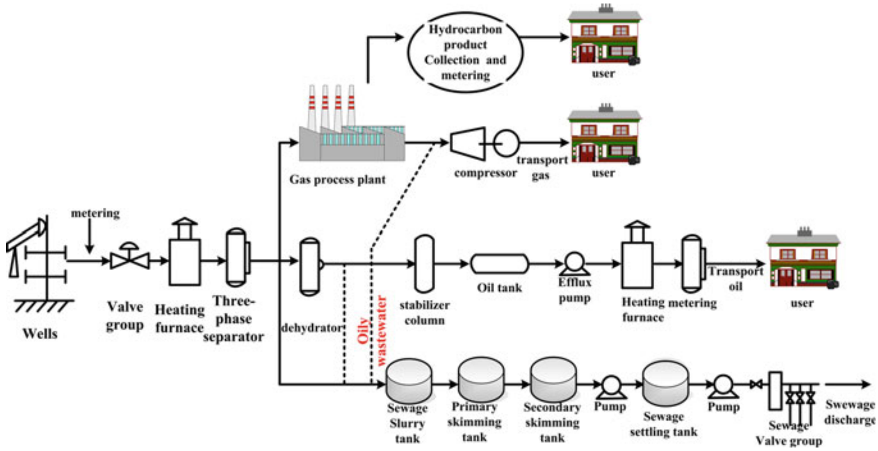


Fig. 51.3 Gathering system process flow

Separated crude oil flows into gas process plant to be stable processing. Finally it will be stored in oil tank and transported to users by pipeline after being processed by electric dehydrator; Gas separated from top of three-phase separator will be transported to gas process plant; Some oily wastewater separated from the bottom of three-phase separator after being settled from electric dehydrator and sewage settling tank down to treatment of primary skimming tank and secondary skimming tank will be injected into oilfield, while the other oily wastewater will be discharged after treatment of biochemical processing station.

In order to create an MFM model of this system, the key step is to answer "what are the goals of the system?", and "what kind of sub-systems to realize the sub-goals?" The overall goal is to treat oil and gas from well and transport them to users or store in tank. The overall goal can be decomposed into several sub-goals and sub-systems are defined below to achieve sub-goals:

- *Metering system*. Goal: Oil, gas, and water produced from wells to be metered to provide dynamic material for oilfield production.
- *Separation of oil, gas and water mixture system*. Goal: Mixture are separated into liquids and gas, and liquids are separated into wet crude and oily sewage, when necessary, to extract solid impurity.
- *Dehydration system of crude oil*. Goal: Make the water content of crude oil conform to standard.
- *Stabilization system of crude oil*. Goal: Make the saturation steam pressure of crude oil conform to standard.
- *Storage system of crude oil*. Goal: Keep the balance between production and sales.

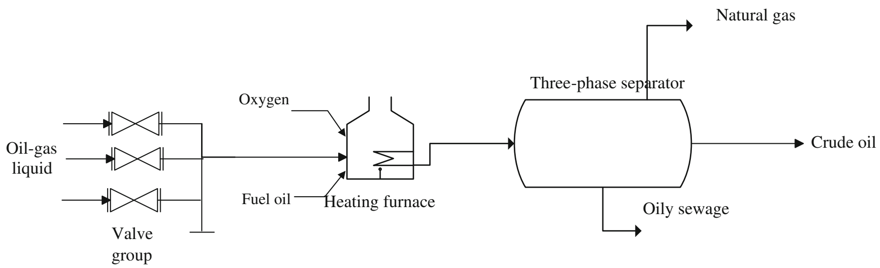


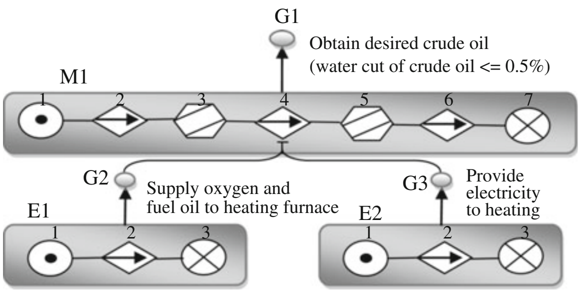
Fig. 51.4 Dehydration system of crude oil

- *Dehydration system of natural gas*. Goal: Dehydrate water from natural gas to prevent generating hydrate.
- *Light hydrocarbons recover system of natural gas*. Goal: Extract hydrocarbon liquid to prevent precipitation.
- *Storage system of hydrocarbon liquids*. Goal: Keep the balance between production and sales of LPG and NGL.
- *Transportation system*. Goal: Transport crude oil, natural gas, LPG and NGL to users.

In the above the gathering system is divided into nine sub-systems, for sampling, dehydration system of crude oil is taken for example, which consists of the piping from wells through valve group to heating furnace and three-phase separator, shown in Fig. 51.4. The oil–gas liquid is heated by heating furnace and the heating furnace needs oxygen and fuel oil in order to work, then the oil–gas liquid is transported by pipeline into three phase separator being separated into natural gas, crude oil, and oily sewage.

The goals of this system are: "Obtain desired crude oil (water cut of crude oil $\leq$0.5 %)," "Supply oxygen and fuel oil to heating furnace," and "Provide electricity to heating furnace."

Fig. 51.5 An MFM model of Dehydration system of crude oil



The functions of the system are, among others, the pipeline's ability to transport liquid, the heating furnace's ability to heat liquid, and the three-phase separator to exact natural gas, crude oil, and oily sewage. An MFM model of this system is shown in Fig. 51.5.

In this MFM model, there are three flows. The network marked "M1" is a model of mass flow, which describes the gas-oil mixture liquid flow from the valve group to the Stabilization system of crude oil. The network marked "E1" is a model of thermal energy flow, which describes the transport of oxygen and electricity to heating furnace. The network marked "E2" is a model of electrical energy flow which describes the providing of electricity to heating furnace. In the network M1 the functions are, from left to right: (1) a source of gas-oil mixture liquid produced from wells; (2) a transport, realized by valve group; (3) a balance, realized by the pipeline between valve and the heating furnace; (4) a transport, realized by the furnace that heat the gas-oil mixture liquid; (5) another balance, realized by the pipeline between the heating furnace to the three-phase separator; (6) another transport, realized by the three-phase separator; (7) a sink of crude oil. The networks E1 and E2 contain energy flow functions describing the flows of thermal, chemical and electricity energy.

It should be noted that MFM describes how different flows enable each other. In this system in Fig. 51.5, it can be seen that the flow of E1 and E2 are necessary to keep the heating furnace running for liquid flow. Such nature can provide possibility for revealing the coupling relationship among failure modes.

51.3 Failure Modes Analysis Based on MFM-HAZOP Model

HAZOP studies are used to investigate deviations from a norm, see Ref. [15]. In MFM, the function nodes values under normal states will be limited in constraint range, however, once the components corresponding function nodes fail, the changes of values will be produced. Therefore based on the MFM reasoning rules [16], function nodes and deviations, it is possible to build a MFM-HAZOP model to identify the failure modes.

In Fig. 51.5, there are three goals of this system. If any goal fails to be achieved, it means one failure mode has been found. The following are failure states of the flow function nodes (Table 51.1).

We assume that the main goal of obtaining the desired crude oil would fail, which means the water cut of crude oil exceeds 0.5 %. Based on the MFM-HAZOP model, the following possible failure modes have been identified (Table 51.2).

Table 51.1 Failure states of the flow function nodes

Function nodes	State	Condition	Description
Source	Low flow	$F < F_{low}$	Explain the flow lower than expectation
	High flow	$F > F_{high}$	Explain the flow higher than expectation
	Impurity	Unexpected mass	Explain source contain unexpected mass
Transport 1 (valve group)	Low flow	$F < F_{low}$	Explain the flow lower than expectation
	High flow	$F > F_{high}$	Explain the flow higher than expectation
Balance (pipe)	Leakage	$\sum F_{in} > \sum F_{out}$	The flow of balance input more than the output flow, flow loss
Transport 2 (heating furnace)	High temperature	$T > T_{high}$	Explain the temperature higher than expectation
	Low temperature	$T < T_{low}$	Explain the temperature lower than expectation
	High pressure	$P > P_{high}$	Explain the pressure higher than expectation
	Low pressure	$P < P_{low}$	Explain the pressure lower than expectation
Transport 3 (three-phase separator)	High temperature	$T > T_{high}$	Explain the temperature higher than expectation
	Low temperature	$T < T_{low}$	Explain the temperature lower than expectation
	High pressure	$P > P_{high}$	Explain the pressure higher than expectation
	Low pressure	$P < P_{low}$	Explain the pressure lower than expectation
	High level	$V > V_{high}$	Explain the level higher than expectation
	Low level	$V < V_{low}$	Explain the level lower than expectation
Sink	Below the index	Water cut of crude oil ≤ 0.5 %	
	Above the index	Water cut of crude oil ≥ 0.5 %	

Table 51.2 Possible failure modes while water cut of crude oil ≥ 0.5 %

Function Node	Deviation	Possible reason	Result	Safety measurement
Heating furnace	High oxygen flow	Fan baffle open widely (human-error/fan damage)	Temperature of fire box and outlet is low, which goes against emulsification and the separation of water and oil.	Oxygen analyzer
				Depression meter
		Seal baffle of fuel open widely		Fan outlet pressure indicator
	Low fuel oil flow	Fuel management system fails		Fuel system pressure indicator
		Turn down the fuel oil valves		Low pressure, alarm of fuel system
	Low fuel oil pressure	Low fuel system pressure		Heating furnace outlet temperature indicator
		Turn down the pressure control valves		Heating furnace outlet and inlet pressure indicator
	Low furnace fire box temperature	The capacity of fire box is big		
		Insufficient oxygen		
		Low fuel flow		
		Low fuel air flow		
		The heated area of furnace pipe is big		
		The absorption ability of furnace pipe is poor		
		Low fuel pressure		
		Fuel oil quality does not meet the requirements		
		Adjustment ability of tempering mouth is poor		
		Low combustor		
		Load		
		Low fuel pressure		
	High atomized steam pressure	High steam system pressure		
		DPCV fails		
	Low outlet pressure	Crude oil flow increases		
		Low crude oil temperature		
		Fuel supply decreases		

(continued)

Table 51.2 (continued)

Function Node	Deviation	Possible reason	Result	Safety measurement
Three-phase separator	High water–oil interfacial	The oil–water interface design is on the high side	Sediment time of oil–water mixture is not enough which leads to higher water cut of crude oil	Content gauge
				Outlet flow
				Sensor separator
		The liquid level controller is out of work, when the oil–water level is high, the level can't be adjusted		Internal temperature sensor
				Pressure gauge
	Poor separator drainage	The pipe jams	Water cut of crude oil is high in oil chamber	
		The outlet valve fails		
	Low separator temperature	The crude oil temperature heated in heating furnace is	Crude oil viscosity increases, which goes against oil–water separating.	
		Low, which leads to lower temperature of crude oil flowing into separator	Finally it leads to high water cut of	
		The heater's fault inside the separator leads to lower heating temperature	Crude oil	
Pipeline	High resistance	Inside the pipeline there is certain thick paraffin precipitation	High resistance is favor of forming emulsion, which leads to high water cut of crude oil	Pipeline pressure sensor
				Pipeline flow sensor
				Paraffin removal by chemical additive or tub cleaner
				Being heated to transport
				Add sedimentation agent to transport
				Blend water to Transport
				Add drag reducer to transport
				Avoid hill-climbing point
		Crude oil contains a certain amount of sand		Clean and purge pipeline
		Low crude oil temperature		
		Low crude oil speed		
		Pipeline inner wall is not smooth enough		
		Crude oil contains a lot of gum and asphaltene.		
		The influence of external terrain		
		Mesh jams		

(continued)

Table 51.2 (continued)

Function Node	Deviation	Possible reason	Result	Safety measurement
Valve group	Low medium pressure	Flow backwards	High water cut of crude oil	Regularly examine valve group to ensure intact components
		Incompletely seal		
		Improper valves		
		Non-sensitive valve		Select proper valve
		Connection and truncation		Change inactive spring in time
		Valve group control system fails		Adjust valve open pressure

51.4 Conclusions

In this article, a failure modes analysis based on MFM-HAZOP in processes with flows of mass and energy has been presented. The MFM-based concept has been tested previously but not on installations as critical as gathering system. This method provides the base of gathering system safety decision to guarantee the gathering system safely operating.

Acknowledgments This project is supported by National Natural Science Foundation of China (Grant No. 51005247); National High Technology Research and Development Program of China (Grant No. 2008AA06Z209); Beijing Nova Program(Grant No. 2010B068); Special Items Fund of Beijing Municipal Commission of Education; Research Funds Provided to New Recruitments of China University of Petroleum-Beijing(Grant No. QD-2010-01) and China University of Petroleum-Beijing.

References

1. Tweeddale M (2003) Failure mode and effect analysis (FMEA) and failure mode and effect criticality analysis (FMECA). In: Managing risk and reliability of process plants. Elsevier Inc, Burlington, p 468–472
2. Dussalut HB (1984) Automated FMEA—status and future. In: Proceedings of the annual reliability and maintainability symposium, p 1–5
3. Atlee J (2006) Software engineering theory and practice. Pearson Prentice Hall, New Jersey, pp 264–266
4. Batzias AF, Batzias FA (2004) Fuzzy fault tree analysis as a means for computer aided technology transfer to small/medium anodizers. Comput Aided Chem Eng 18:577–582
5. Modarres M (1993) Functional modeling of complex systems using a GTST-MPLD framework. In: Proceedings of the international workshop on functional modeling of complex technical systems, Ispra, Italy
6. Öhman B (1999) Failure mode analysis using multilevel flow models. In: Proceedings of the 5th European control conference, Karlsruhe, Germany
7. Dahlstrand F (2000) Methods for alarm reduction with multilevel flow models of industrial processes. Ph.D. Thesis, Lund University: Department of Information Technology, Lund
8. Larsson JE (1992) Knowledge-based methods for control systems. Ph.D. Thesis, Lund Institute of Technology: Department of Automatic Control, Lund
9. Lind M (1994) Modeling goals and functions of complex industrial plants. Appl Artif Intell 8:259–284
10. Lind M (1990) Representing goals and functions of complex systems-an introduction to multilevel flow modeling. Institute of Automatic Control Systems, Technical University of Denmark, Lyngby, report no: 90-D-38
11. Dahlstrand F (1998) Alarm analysis with fuzzy logic and multilevel flow models. In: Proceedings of the 18th annual international conference of the british computer society special group on expert systems, Cambridge, England, p 173–188
12. Larsson JE (1996) Diagnosis based on explicit means-end models. Artif Intell 80(1):29–93
13. Öhman B (2001) Real-time diagnosis of industrial processes using multilevel flow models. Ph.D. Thesis, Lund University: Department of Information Technology, Lund

14. Öhman B (2002) Discrete sensor validation with multilevel flow models. IEEE Intell Syst 17(3):55–61
15. Rossing NL, Lind M, Jensen N, Jørgensen SB (2010) A functional HAZOP methodology. Comput Chem Eng 34(2):244–253
16. Dahlstrand F (2002) Consequence analysis theories for alarm analysis. Knowl-Based Syst 15(1):27–36

Chapter 52
A Review of Machinery Diagnostics and Prognostics Implemented on a Centrifugal Pump

K. K. McKee, G. L. Forbes, I. Mazhar, R. Entwistle and I. Howard

Abstract Centrifugal pumps are a widely used machine found in industries such as water, sewerage, oil, and gas. As a result, it is vital that these pumps are monitored, diagnosed, maintained, or replaced prior to the pump failing to reduce downtime, material, and labour costs. Most companies employ a run-to-fail method or a time-based maintenance strategy to service their pumps, instead of condition based maintenance or a predictive maintenance strategy. This paper reviews the state of art in diagnostics and prognostics pertaining to centrifugal pumps. Attention is given to detailing the methods of application, detection of fault modes and results used by researchers in the main areas of diagnostics and prognostics.

52.1 Introduction

Centrifugal pumps consist of two main parts: the impeller and the casing, also known as the volute. The impeller is a set of vanes, sometimes enclosed, which are used to direct the fluid entering the pump away from the inlet openings. The

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

G. L. Forbes
e-mail: gareth.forbes@student.unsw.edu.au

I. Mazhar
e-mail: i.mazhar@curtin.edu.au

R. Entwistle
e-mail: r.entwistle@exchange.curtin.edu.au

I. Howard
e-mail: ian.howard@unsw.edu.au

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_52,

impeller is connected to a motor and rotates as the fluid is entering the pump. As a result, it transfers kinetic energy from the motor to the fluid. The fluid is then funnelled by the volute in a desired direction, thus changing the kinetic energy to pressure [1].

Current maintenance strategies implemented by many companies are either a run-to-fail method or preventative maintenance method, also known as time-based maintenance (TBM). The problem associated with the former is the cost of time and money that is lost when the pump fails during peak times of use. In order to circumvent this, some companies provide redundant systems which are run either simultaneously or are turned on only when the primary pump fails. Although this may ensure the continuity of the process, in the long run, it is a costly solution since the failed pump's life could have been extended and repairs could have been less costly if they were caught prior to failure. Preventative maintenance methods are an attempt to decrease the off-line time that a machine needs to be maintained by providing servicing on a time-cycle basis. This may mean that parts of the pump are replaced, not because they are about to fail, but because their scheduled life has been used up. As a result, parts which still have a remaining useful life are sometimes replaced, which provides a non-optimal method of maintenance based on cost for parts, labor, and time. In addition, due to unforeseen circumstances, parts of the pump may fail in between servicing times, resulting in possible catastrophic events [2].

Two other maintenance procedures that are not as widely used are condition based maintenance (CBM) and predictive maintenance. Since it has been reported that 99 % of mechanical failures are preceded by noticeable indicators, these two methods utilize these indicators to determine the current state of the machine, and for the latter, to attempt to predict the future state of the machine within a given time frame. Due to the advent of faster processing computers, remote sensors, and database software, computerized maintenance procedures which utilize expert knowledge as well as algorithmic decisions are becoming more feasible for use in industrial settings [2].

Data collected in a CBM program can be classified into two main types: event data and condition monitoring data. Event data pertains to what occurred, such as breakdown and overhaul and their causes, and what was done to the machine, such as installation and maintenance. Condition monitoring data are the information that is related to the health condition of the machine. Both types of data are equally important for a CBM program [3].

Predictive maintenance utilizes prognostic algorithms to determine the remaining useful life (RUL) of the machine. To do so, it deals with fault and degradation prediction prior to them occurring, utilizing statistical data on worst-case failures relevant to usage and time in service. This method is still in its infancy, and thus does not seem to yet be utilized widely in industry [4]. Currently their exits a number of papers which discuss CBM and predictive maintenance holistically, across a broad range of industry sectors. For instance, Jardine et al. [3] and Peng et al. [5] wrote review papers detailing the state of the art in CBM and prognostic technology. Heng et al. [2] wrote a review paper detailing the state

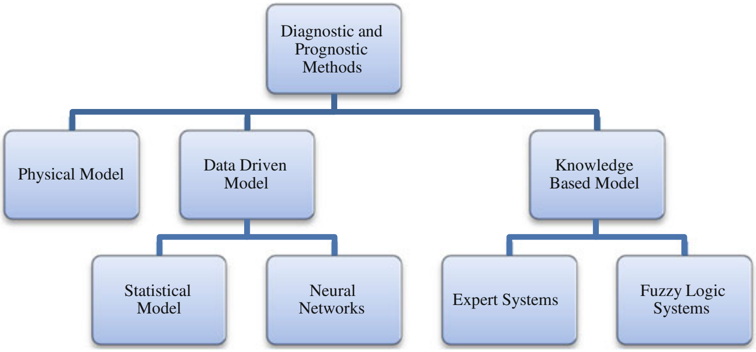


Fig. 52.1 Overall layout of diagnostic and prognostic methods

of the art of prognostics, and their challenges when it is applied to rotating machinery. These papers talk about the CBM methods in general.

Methods for diagnostics and prognostics can be divided into 3 main groups: Physical models, Data driven models and Knowledge based models, c.f. Fig. 52.1. Data driven models are further broken down into statistical models and neural networks. Knowledge based models are broken down into expert systems and fuzzy systems. In addition, some methods utilize a combination of the different models to attempt to exploit their strengths to produce a better diagnostic or prognostic tool. These models will be discussed in general and their implementation by various researchers in detail.

This paper focuses on the current state of research and the application of CBM and prognostics to a centrifugal pump. It will focus more on providing details on how various researchers have attempted to apply the state of the art to the pump, rather than in depth explanation of the methods themselves. To obtain the latter, readers should consult any of the previous referenced review papers. Fault modes of centrifugal pumps have been discussed in detail by McKee et al. [6].

52.2 Instrumentation for Detection

Performing CBM on a centrifugal pump entails determining the condition of the pump using non-intrusive sensor measurements, in essence all CBM methods use some form of sensor gathered data and interpret and analyse this data to achieve the condition and/or predicted life of the machinery. There exist off-the-shelf instruments, as well as standard methods, that are utilized to determine the current state of the centrifugal pump. Some of the instrumentation and measured parameters are as listed:

- Thermometric testing, in which temperature readings are taken upstream and downstream of the pump, has been used to locate problems that cause changes in the temperature of the fluid or machine parts. This is done by placing two test tappings, one on either side of the pump. Changes in temperature are surmised to result from inefficiencies in the pump being converted to heat, which would be picked up by the temperature difference. Any type of recirculation, whether inlet or outlet, must be eliminated and the tests are not applicable to very low or zero flow. Bearing damage may show up in the temperature rise of the bearing itself, but not in the temperature rise of the fluid. The same occurs for losses in seals. This method would show promise provided that steady state conditions can be arranged prior to taking readings [7].

- Thermography, which is a similar type of testing, utilizes a special camera to scan the surface of the pump and the surrounding pipes to create a temperature profile in a color scale. Based on the same principles as thermometric testing, thermography can also capture bearing, misalignment, and seal problems that result in changes in the temperature of the object itself, but were not able to appear in the thermometric testing [7].
- Non-contact proximity displacement probes are ideal for determining the orbit of the shaft, from which it is able to confirm alignment issues, and frequencies present in the system. However, due to mounting issues such as limited access to shafts, these types of sensors are rarely used for pump investigations. Instead, tri-axial accelerometers are usually mounted on pumps to determine their vibrations. FFT analysers are used in conjunction with the sensors, to remove random noise and extract the frequencies found in the signal. Electronic data collectors/analysers are available for field use from several manufacturers, coming equipped with LCD screens or adapting PDAs with sensors to obtain the desired data. Beebe details the correlation between certain vibration frequencies and fault modes that occur in a centrifugal pump [7].
- Visual inspection can be performed using an optical boroscope, an optical fibrescope or a television (digital) fibrescope. Optical boroscopes provide good clarity but are unable to view objects at an angle. Optical fibrescopes are composed as a bundle of glass fibres which provide the ability to view around corners at the cost of a less clear picture than the boroscope. Digital fibrescopes use a tiny TV camera to display the picture on a monitor [7].
- Ultrasonic methods can be used to locate cracking in pump components and map their size and orientation. It is also used to check the thickness of linings of pipes surrounding the pump. Significant skill is required to perform this type of inspection [7].
- Lubrication analysis has been a time-tested method in CBM. Particles in lubricants can describe the wearing process occurring within a pump. The wearing of lubricated surfaces causes tiny particles of material to wear off and be carried in the lubricants and the fluid being pumped. Normal wear and tear on a pump can result in particles up to the size of 8 micrometres in diameter to appear. Analysis of the lubrication also involves determining the amount of particles present, usually reported as in parts per million (i.e., ppm). Abnormal wear results in larger particles created, which are at least 50 micrometres or greater. The shape of the wear debris particles is also useful in the diagnosis [7].

- Motor current signature analysis is the most widely used method for motor condition monitoring. Analysing motor current can reveal damage to the rotor bars, which appear as side bands in the motor current spectrum when the load is above 50 % of its maximum. This damage can also be shown in vibration analysis. Recommended settings for motor current signature analysis are in the frequency range of 0–80 Hz. In this case, the rotor speed must be known so as to get the slip frequency (ac frequency minus the rotor frequency) and the side bands [7].
- A fiber-optic pressure sensor utilizing a Fabry–Perot interferometer was created by FFPI Industries, which is able to obtain the pressure readings of a fluid that has temperatures in excess of 700 °F. Changes in fluid pressure cause slight strains in the fibre, which then leads to a large fractional change in the reflected optical power. This new sensor was tested on a 125 hp double-suction centrifugal pump running at 3,600 rpm [8].
- Wireless vibration sensors can be used in order to acquire information from hard to reach areas on machines. To utilize the wireless sensors for CBM, the overall system must contain a sensor network array, wireless transmitter with an analogue-to-digital converter, wireless access point, data acquisition computer with signal–conditioning capability, software module with fault diagnostics and pattern recognition, data archival system, a database, and a smart diagnostic software module. The wireless technology allows the ability for continuous monitoring capability of centrifugal pumps. This type of system has been implemented at the High Flux Isotope Reactor at the Oak Ridge National Laboratory [4].
- Microphones have been utilized to detect cavitation and other hydraulic problems, which are known to create high amplitude acoustic vibrations that can easily be recorded and analysed to determine its frequencies. As a result, research has been performed in determining which acoustic frequencies are present at the onset and during certain hydraulic problems, mainly cavitation, and if these frequencies are unique so as to be used as indicators for the fault mode.
- An acoustic emission (AE) is the transient elastic wave that is generated by the strain energy that is caused from the deformation or damage within or on the surface of a material. Sources of AE include impacting, cyclic fatigue, friction, turbulence, material loss, cavitation, leakage, etc. It can offer the advantage of earlier defect/failure detection compared to vibration analysis. The limitations of this method lie in the processing, interpreting, and classifying the information from the acquired data. Mba et al. have written an extensive literature review of this topic, with a section dedicated to AE and cavitation in pumps [9].

Some research has been done in the field of acoustics, using microphones, to determine the utility of this method for detecting cavitation. Cudina used acoustics to determine the onset of cavitation in a centrifugal pump. It was found that there is a distinct increase in the amplitude of half of the blade passing frequency when cavitation begins [10].

Al Thobiani et al. used conventional vibro-acoustic measurements to create different dimensionless parameters that could be utilized for cavitation diagnosis on a centrifugal pump. Two identical accelerometers with a bandwidth of 20 Hz–10 kHz were placed on the pump casing at the suction and discharge side of the pump. It was found that the suction side accelerometer was less sensitive to cavitation than the discharge side. The acoustics of the pump was measured using a microphone for airborne noise and a hydrophone for fluid acoustics. Each transducer has a bandwidth of 20 Hz–20 kHz. It was found that vibration trends of both peak factor and kurtosis showed a gradual decreasing trend as cavitation occurred. In addition, spectral entropy, which is a measurement of the disorder in the system, was used. In its normalized form, it produces a value close to 1 when its amplitude distribution is flat, such as when amplitudes of each frequency component are equal. It also produces small values, close to 0, if amplitudes in the frequency spectrum are concentrated in a few frequency components. Hence, in essence, it is capable of evaluating the vibration signal's spectral structure. It was found that acoustic trends are more consistent with the cavitation progression, while vibration trends indicate the start of the cavitation too early. This may be due to the fact that acoustic measurements can capture more global dynamics of the pump than surface vibrations obtained at a local position [11].

52.3 Physical Model Based Systems

Physical model-based systems employ mathematical models to describe the physics of the system and their failure modes, such as crack propagation and bearing failure. Many physical systems can be modeled simply using masses, springs, dampers, and distributed and point loads. For example, a simple physical model of car may depict a car as being a body of mass suspended on four spring-damper systems, resulting in second order differential equations to describe the system. Crack and fault mode analysis may be implemented using finite element methods or alterations in the differential equations for the system.

These models are developed by system experts, and are validated using a large data set from the machine. This type of approach uses residuals as features. Residuals are the differences in values between the output of the real system based on sensor data and the outputs of the mathematical model. Under normal conditions, it is assumed that the system will operate within certain limits, predetermined by manufacturer's tolerances, statistical techniques, and by a system's analysis. When the residuals exceed predefined thresholds, malfunctions are assumed to exist in the system. These types of models are configured specifically for the system being monitored. The limitations to this type of system are their high cost and component speciality. In addition, it is difficult to build a model that incorporates all aspects of a system [5].

Kallesoe et al. devised a nonlinear model of a centrifugal pump, which was independent of the operating point in which the pump was running. The inputs to the pump were the shaft torque, shaft speed, pressure produced by the pump and the volume flow through the pump. The five faults considered were: clogging inside the pump, increased friction due to either rub impact or bearing faults, increased leakage flow, performance degradation due to cavitation, and dry running. This model was tested against a Grundfos 1.5 kW CR5–10 pump, and was shown to distinguish four out of five faults, but was still sensitive to the operating point despite its non-linearity [12].

Kallesoe et al. also created a model based fault detection algorithm that utilized only the electrical measurements on the motor and the pressure delivered by the pump. The centrifugal pump system was divided into two subsystems—the electrical and mechanical/hydraulic components. The only variables that were measured were the stator voltages and currents of the motor and the pressure produced by the pump. Four faults were considered for this system: clogging inside the pump, increased friction due to either rub impact or bearing faults, performance degradation due to cavitation, and dry running. The first two were caused by impurities in the liquid and wear, the third caused by too low inlet pressure, and the last caused by faults in the surrounding system. The faults in the system were detected using an Analytical Redundancy Relation (ARR), which is a relation that links a subset of selected variables of the system. If the links were violated, a large residual was calculated that reflects the discrepancy between the actual behavior of the system and the expected behavior given by the model, and a fault is said to have occurred in the system. The method was once again performed on a Grundfos 1.5 kW CR5-10 pump. The method was able to detect the faults placed into the system, however, transient phases in the system created problems for the algorithm since there was an assumption that the pump was running at a constant speed [13].

Isermann surveyed fault detection using modeling and estimation methods. In essence, there are measurable and unmeasurable, but estimated, variables in a model, which have certain tolerances of the normal value. The model used sensor data to monitor necessary values within the system. A fault message was triggered if variables were found to be outside their tolerance. Then a fault diagnosis (location and cause of the fault) was established, followed by fault evaluation (an assessment of how the fault will affect the process). In the end, a decision must be made as to how to handle the fault—whether to deem it tolerable or eliminate the cause of it. A prediction method to predict the time of failure, or exceeding a threshold, was desirable. The method was implemented on a centrifugal pump with a water circulation system, driven by a speed-controlled direct current motor. The problem with the implementation of this method was that the system tracked a leak as if it was not a problem in the system, but rather a normal change to the system. Hence, instead of being insensitive to process noise but sensitive to leaks, it processed the leak information which caused the fault not to be detected [14].

Nordmann et al. created a system model of a rotor using a mass, damping, stiffness, and forcing function matrices. Active magnetic bearings (AMB) were used on the centrifugal pump, from which the inputs, outputs and system characteristics were able to be measured. The pump data that was used included rotational speed, flow rate, head, radial clearance of the piston seal, and the radial clearance of the impeller seal. From the magnetic bearings, the air gap, AMB force, pre-magnetisation current, windings per pole pair, and the cross sectional area of the pole were used. From this information, the magnetic bearing force was calculated. Twenty-five measureable transfer functions relating the 5 inputs with the 5 linear independent force excitation patterns were created, which served as references for the response of the system. Huge deviations from the model parameters indicated a fault in the system. Identification of the acting fault was done by evaluating the symptom-fault relations or by creating faults in the model parameters and comparing the faulty-model parameters with the actual data. The method was validated on a centrifugal pump which had no fault, and another with a dry-run [15].

52.4 Data Driven Based Systems

Data driven methodology is based on statistical and learning techniques. Data driven systems do not require knowledge of the underlying system. They track patterns or changes in the measured data and correlate these patterns to different fault conditions. These types of systems can be classified into two different categories: statistical approaches and AI approaches. Statistical approaches utilize multivariate statistical methods, state space models, and regressive models. An example would be determining the possibility that an impeller in a centrifugal pump would fail based on the amount of time the part has been in use and a probability density function of the failure times of impellers.

AI approaches usually implement artificial neural networks (ANN) and its variants. ANNs consist of three types of layers: input layer, hidden layer, and output layer. Each layer contains a number of mathematical elements, called "neurons" that receive a weighted sum from all nodes in the previous layer, and output a numerical answer based on the function governing the neuron. ANNs is a mathematical attempt to model the reasoning and pattern matching capability of the human brain. There are two main types of training methodologies: supervised training where the ANN is trained using a set of inputs and a pre-classified output, and unsupervised training where the network is used to classify network inputs. An example of supervised training is providing the ANN with various vibrations of a pump, having each example associated with a fault mode. While the neural network is being trained, it attempts to create weights and biases between the neurons so that it can distinguish the different fault modes based on its inputs. If trained properly, a random set of vibrations would cause the ANN to state the correct associated fault mode. Unsupervised learning is utilized when there is no clear understanding of what is found in the data, and it is desired for the ANN to separate the data based on the most distinct characteristics. An example of this would be the reorganization of the clothes in a closet. Suppose the clothes in a closet need to be separated into groups, and various traits of the clothes are provided such as size, color, type of clothes, style of clothes, manufacturer, material, and season that it is worn in. All this information would be placed into the unsupervised ANN, which will determine which characteristics are most distinct, thus resulting in the best separated groups. The clothes could then be arranged according to the groups set up by the ANN.

52.4.1 Statistical Approaches

Cempel discussed the application of vibroacoustical diagnostics to a rotating machine, with application to a centrifugal pump, and outlines a possible method. The author first suggested to maximize the signal to noise ratio in the prefiltering of the signal. This is to ensure that the analysis would be done on the vibrations of the machine, and not on those of the surrounding components. Otherwise, the accuracy of the analysis is compromised. Frequency analysis (such as a Fast Fourier Transform or a cepstrum analysis) would then be used to process the signal for useful information about the system and to differentiate between various faults (such as bearing instability, imbalance, and non-alignment). This would create a signal estimate or measure which acts as an indicator for the condition of the machine. Finally, this culminated with a method forecasting remaining life [16].

Sinha et al. performed diagnostics on one of India's main coolant centrifugal pumps (MCP-3). Nine accelerometers were placed on the MCP-3 in order to extract data to perform modal testing, as well as creating a higher order spectra (HOS) and wavelet transform (WT) analysis. When performing the WT analysis, it is necessary to reliably determine the proper wavelet to use. The modal data (natural frequencies, mode shapes, and damping) were extracted using multi-degree curve fitting on the experimentally derived frequency response functions (FRFs). The HOS bi-spectrum and tri-spectrum were used for the detection of higher harmonics and their phase relationship in the signal, as well as detecting behaviour in structural dynamic response, such as the detection of transient signals, looseness at the support, and nonlinearity in the system's stiffness. Experimentally identified natural frequencies, and causes thereof, were found to be the roots of the failure. Solutions were then implemented to remove the problems [17].

Lee et al. fitted an inducer at the inlet of a centrifugal pump, where pressure fluctuations that cause cavitation in the centrifugal pump were measured at the inlet and outlet locations of the inducer. The time–frequency characteristics were analysed using a Choi-Williams wavelet distribution. It was found that cavitation causes higher frequencies to appear at the exit end of the inducer [18].

Zhang et al. utilized an eigenvector analysis method, Principal Component Analysis (PCA), to select and combine several vibration and feature frequencies to obtain the best predictors of the condition of the machine. A key issue with using PCA is determining how many principal components are necessary to uniquely identify the fault modes present. Diagrams were created using 2 and 3 principal components to visualize the data. The data was assumed to follow a joint multivariate Gaussian distribution, and a 95 % confidence level was used to detect deviations from the norm. Data was extracted from a double-suction centrifugal pump [19].

52.4.2 Neural Networks

Heng et al. propose a feed-forward neural network (FFNN) that has d+1 inputs, where d is the number of previous time intervals of recorded data, and h outputs, where h is the number of time intervals that the network is desired to predict ahead. The FFNN training targets were computed using an adapted Kaplan–Meier estimator and a degradation based failure probability density function to help predict a non-linear time to failure, instead of a linear time to reach a cut-off value. The model was tested using the data from Gould 3175L centrifugal pumps at Irving Pump and Paper, which have a high incidence of unpredicted failures. Vibration signals were collected at 8 locations on the pump, and then pre-processed into 5 frequency bands, an overall summary of the 5 bands, and an acceleration value. This was then used to compute a "current asset condition value". 10 delayed values and 1 current value were inputted into the FFNN, which had 15 hidden nodes, and 5 output nodes which predicted 5 intervals ahead. The training function used was a gradient descent algorithm with momentum back propagation to prevent the network from getting stuck in shallow local minimums. The method showed the nonlinear trend in failure and a reasonable probability of failure in the predicted intervals [20].

Tsai et al. attempted to determine the best type of neural network to use in the diagnosis of a mechanical pumping system. The three types of neural networks tested were the back propagation (BPN), radial basis function (RBF), and adaptive linear (ADALINE) neural networks. The 8 parameters for the diagnosis of the pumping system included the power factor, voltage, current, power, rpm of motor, degree of vacuum, pressure at pump outlet, and flow rate. Ten sampling conditions, one normal and nine fault conditions, with eight sampling channels, were considered. Results showed that the best neural network for online diagnosis was BPN, followed by ADALINE, then RBF. However, if the mechanical system is diagnosed off-line, then RBF was shown to be the best one to use [21].

Rajakarunakaran et al. stated that a desirable diagnostic system possesses quick detection and diagnosis, isolability, robustness, novel identifiability, classification error estimation, adaptability, explanation facility, modelling requirements, storage and computational requirements, and multiple fault identifiability. Neural networks are ideal to handle fault detection and diagnosis problems due to their non-linear function approximation and adaptive learning capabilities, however, they need a lot of data to develop the network before being used in real-time applications. Adaptive resonance theory (ART) refers to a class of self-organizing neural architectures that allow plasticity to learn new patterns while maintaining previously learned patterns. The ART network was tested on a centrifugal pump system, which has sensors reading the voltage, ammeter, vacuum pressure, speed, time, the amount of metre rise in the collecting tanks, and the time for the number of revolutions in the energy meter. The ART was trained one fault at a time. Sensor data was normalized to give equal priority to all the inputs. The input data contained 11 features, and the output contained a vector of 20 outputs, each output signifying a different fault. A neural network with the ART1 algorithm was compared against a neural network with a back propagation algorithm. The ART1 method was able to operate 4 times faster than back propagation, and classified the testing data with 100 % accuracy, compared to the 99.3 % accuracy of the back propagation algorithm [22].

Saxena et al. utilized genetic algorithms (GA) to maximize the weights and biases of an ANN. The GA-evolved ANN outperformed the standalone ANN, sometimes with 77 % better results, when tested for bearing diagnostics [23].

Ugechi et al. utilized an ANN with three hidden layers to perform diagnostics on a centrifugal pump. The inputs were vibration-velocity amplitude, motor power, equipment frequency, and spike energy. Faults that were considered included misalignment, imbalance, bent shaft, mechanical looseness, and poor bearing condition. The program was tested on a system containing a misalignment and a defective bearing. The output of the program showed the input sensor information, manifestation, fault classification, and instructions to remedy the situation. The method was tested on a Giabbioneta 36.5 kW centrifugal pump [24].

52.5 Knowledge Based Systems

Due to the difficulty of building a mathematical model of a physical system that has real-world applications, some research leans towards knowledge-based methodology, which requires no physical model to be constructed. Two typical examples of this are expert systems and fuzzy logic systems.

Expert Systems have been used since the 1960s to solve problems that were usually solved by human specialists. It can be considered as a computer system programmed to exhibit expert knowledge in a particular domain. Rules are expressed using IF–THEN statements or implemented using decision trees, and attempt to cover all possible scenarios dealing with fault modes of a system. However, the expert system cannot handle new systems that are not covered in its knowledge base, and thus is limited in its scope, and confined to the knowledge of the expert who programmed it.

Fuzzy logic attempts to arrive at a definite conclusion based on vague, imprecise, noisy, ambiguous or missing information. It models system behaviour on a continuous number scale in mathematics, rather than mapping to discrete values. In addition, it utilizes qualitative and imprecise reasoning statements as a rule base in order to produce a simpler, intuitive, and better behaved model. Fuzzy logic systems are usually incorporated with other techniques such as expert systems and ANNs.

52.5.1 Expert Systems

Isermann creates a methodology for fault detection in machines. The first step was to input information that is based on physical laws, knowledge provided by an engineer, or other processes. This information was then fed into a knowledge base that was created using analytical models, such as mathematical models showing its static and dynamic processes, and heuristic models, such as those using fault trees, process history and fault statistics. Static process models were expressed as a nonlinear polynomial, while dynamic processes were linearized around one operating point and represented as an ordinary linear differential equation. After the selection of an analytic or heuristic model, rules created were based on the symptoms being provided to determine which fault was present, if any. For a DC motor and a centrifugal pump setup, the measured signals were the voltage, armature current and speed of the DC motor, input and output pressure, and mass flow of the pump [25].

Ebersbach et al. created an expert system designed to utilize 75 rules that handled 54 different machine faults common to mechanical systems found in most fixed plants. The machine components that were addressed were roller and journal bearings, spur gears, belt drives, couplings, and centrifugal pumps. Rules were developed by relating the physical fault condition to the frequencies emitted by the machine, and analyzing the vibration data for unique fault signatures, such as high amplitude peaks at certain frequencies. These rules were implemented using If–Then statements in Microsoft Visual Basic Code, due to the easy implementation of a user interface. The peak detection algorithm detected peaks by first normalizing peak amplitudes based on the largest peak in the spectrum, and then placing threshold values, also known as alarm limits, on the normalized value. The benefit of this method is its applicability to the analysis of a machine that lacks historical and/or manufacturer data. In addition, a quantitative confidence factor ranging from 0 to 1 was included in the algorithm, which was dependent on the frequency and amplitude of the fault frequency compared to the theoretical frequency and absolute value of the amplitude. A value of 0.5 signified that the fault threshold has been breached. The system was tested on a spur gear test rig and was able to detect a bent output shaft, overload, and contamination successfully [26].

Parrondo et al. created an expert system that was dependent on the relative amplitudes of four different frequencies: the frequency of the impeller's rotation, the first and second harmonics, and the blade-passing frequency. Amplitudes were divided by the corresponding value under normal operation with the best efficiency flow rate. If the ratio was greater than 1.1, then a value of "+1" was assigned to that frequency. Similarly, if the ratio was found to less than 0.9, then a value of "$-1''$ was assigned. Deviations of less than 10 % were assigned a "0" to signify their insignificance along with a "+" or "$-$" to state whether it was an increment or a reduction in value from the reference. In addition, a difference of 2 millibar for pressure signals and 0.1 m/s^2 for acceleration signals was considered the minimal amount of deviation from the normal value to be considered significant. Six abnormal situations (flow rate greater than the best efficiency flow rate, lower flow rate, cavitation partially developed, cavitation fully developed, presence of an obstacle in the inlet conduit, rotational speed greater or lower than the specified value) were able to be uniquely identified based on the pattern produced by assigning these values to the four different frequencies. Acceleration on various parts of the pump was not used since it was thought that the same acceleration can be caused by multiple parts of the pump, and thus can't be used in order to uniquely determine a problem with the pump [27].

Li et al. produced an expert system driven by data mining the sensor data to find certain conditions that could determine the onset of a fault. Rules were created with the use of rough sets, which eliminated redundancies and information that does not help to determine faults in the system. Initially starting with all the sensor outputs, the data was placed into a table. Then the data was discretized into one of five different categories, ranging numerically from −2 to 2. The data was scanned to see if any fields were missing from any sensors. If there were, then the fields were filled with all possible outputs, ranging from −2 to 2, that the sensor can obtain. For example, if the sensor has the possible values of −2, 1, and 2, then three lines were created for this sensor, each containing a different possible output. Then applying a rough sets approach to creating rules, a minimal amount of sensors were selected to determine if a fault was present in the system. The set of rules were created, which combined several sets of sensor outputs, to create the minimal amount of rules needed to diagnose the system. Once the system was trained, subsequent acquired data was checked for new information, such as a new sensor, that has not been taken into account. If there was no new data, the acquired data was then discretized and the rules were applied. The system was tested on two parallel centrifugal pumps connected with valves and piping to 2 storage tanks. The sensor data that was obtained were the flow rate, pressure at the suction and discharge of the pump, bearing housing temperature, bearing vibration in both the horizontal and vertical direction, and the seal oil flush temperature. Rules were created based on all the data from the sensors, except for one which was used to test the rules. The tested set of information was classified correctly [28].

Kollmar et al. created a machine learning algorithm that utilised the algorithm See5 to create its decision trees. It was claimed that on average, decision trees outperform the neural network at its classification rate, having a training time which is 2–3 orders of magnitude shorter compared to neural networks. A cost matrix was produced in order to relate the level of classification of a predicted class to the actual class, weighing misclassifications higher than those classified correctly. Entries were placed on a scale of 0–3, 0 being correct classification and 3 being total misclassification. The pump was equipped with different sensors, which were able to generate up to 40 features that were preselected by a human expert in order to reduce the amount of data. This was compared to the performance of various wrappers, genetic algorithms, and filters. Wrappers, auxiliary algorithms used in conjunction with the basic machine learning algorithm, were used to aid in classification of the data. Four different types of wrappers were used to classify the data: Best Features (BF), Sequential Forward Selection (SFS), Sequential Backward Selection (SBS), and Branch and Bound (BB). For pump data, BF and BB were considered unsuitable for the pump data, and hence, only SBS and SFS have been examined. Genetic algorithms were explored as a wrapper to aid in the clustering of data. Filters were used to determine optimal subsets, regardless of the machine learning algorithm. Three groups of filters can be discerned: Wrappers combined with a fast machine learning algorithm, such as

Nearest Neighbor; blackbox algorithms which were standalone algorithms to determine the suitable feature subset; and filters that utilized the class topology in the feature space, such as the use of various distance functions or classification based on density functions of the classes. In the end, a single stage SFS wrapper with See5 proved to outperform almost all of its competitors, having Bhattacharyya distance come in as a close second [29].

Ebersach et al. created an expert system that utilized two common techniques for fault detection and tracking, namely analysing vibration data and analysing oil and wear debris. The expert system that was devised consists of three parts: VES (a vibration analysis expert system), OWDES (a oil and wear debris expert system), and CES (a combined analysis expert system). Information concerning the setup of the system was initially entered into the system so as to calculate fault frequencies, as in the case for VES, and for OWDES to perform an elemental analysis. VES was created with the capability of analyzing and interpreting tri-axial vibration spectra, time domain, and demodulated spectra data. The knowledge base was developed to detect faults in common industry machinery such as journal bearings, roller bearings, spur gears, couplings, belt drives, and centrifugal pumps. The algorithm used was able to detect a total of 54 faults using 75 rules. In addition, a confidence factor was created to describe the severity of the identified fault. This consisted of the sum of two factors: a frequency factor and an amplitude factor. The factors were calculated by comparing the actual and theoretical values of the fault frequencies, and the amplitude levels in the vibration spectra to the threshold levels, then using fuzzy logic to obtain a value between 0 and 1. These two values were then scaled to be between 0 and 0.5, then summed to produce a value between 0 and 1. OWDES required the input of wear particle types and relative concentrations, and particle color and relative concentrations. It was optional to also input the particle count, viscosity, water concentration, elemental analysis of wear particles and contaminants, chemical index, and total acid number. Once the information was inputted, the particles were then categorised using a previously developed artificially intelligent wear particle identification algorithm that was applied to images using a laser scanning confocal microscope. There were 14 particle types that were associated with wear modes, among them were cutting, fatigue, rubbing, sliding, adhesive, and a combined case of sliding and adhesive wear. The severity of the fault was determined using a confidence factor, which was calculated using fuzzy logic by comparing the concentration of the wear particle with those that were previously assigned as 'Fault' and 'Severe' thresholds. The faults that can be detected using oil analysis are contamination, water in oil, use of the wrong lubricant, oil thickening and dilution, oil additive depletion, and oxidation. The sorting algorithm required knowledge of all various machine components and their elemental compositions. Analysis involved sorting components which contained the greatest number of elements in the lubrication in critical concentrations to the lowest. The next step was CES, which utilized the other two expert systems (OWDES and VES) and categorized their findings into fault groups. A root cause analysis was also performed, where faults were classified according to primary, secondary and unknown chronological faults, and identified according to regions in the machine. The remaining life algorithm was based on the tribology wear equations. The system was tested on two types of gearboxes. The detected faults were 2 bearing looseness, 2 bearing faults, gear operating fault, bent shaft, and gear fatigue [30].

Garcia et al. created an expert system called PUMPX to diagnose problems with centrifugal pumps. Using a series of questions, this knowledge-based program determined the most likely cause of the problem found in the system. Conclusions were derived from following a series of case-switch and if–then statements [31].

Lihovd et al. created a program called ROMEX, which is a rule-based fault diagnostic system for rotating machinery. The user inputs data about the system to configure the program to the user's needs. Eighteen faults were programmed for the centrifugal pump. It utilized "Certainty Addition". To do so, it first calculated a Symptom Strength, which rated the symptom of the problem on a scale from 0 to 100. This was multiplied by a weight given to the symptom for a particular fault type, and then divided by 100 to obtain the actual contribution of the symptom to a specific fault. Finally, a "Fault Probability number" was calculated, which helped rank the possible faults for a given symptom. No validation was provided [32].

Yang et al. created an expert system called VIBEX (VIBration Expert) to aid in the diagnosing of abnormal vibrations for rotating machinery. It utilized two parts: a decision tree (VIBEX-DT) and a decision table (VIBEX-TBL). A decision table was created between causes of faults and symptoms from empirical knowledge obtained from an expert in the field. This information was placed in the program as a set of rules in the form of IF (symptom) THEN (cause). The users were able to select the information based on largest frequency, direction and location of the largest amplitude, and the amplitude response during start up and shutdown. A Bayesian algorithm was then used to obtain the confidence factors (cf), which were used as probabilities of the causes of the vibrations. The decision tree (DT) portion of the expert system utilized the algorithm C4.5. The DT required defined classes that represent the vibration causes, and attributes that represented the vibration phenomena that were grouped into sets for the machine learning process. When data required by the DT were not known, probabilistic results were used to obtain the unknown attributes. Information-entropy evaluation function, which was based on information theory, was used as the selection criteria for the test set on which the DT was trained on. For testing purposes, the DT was trained with 14 classes, based on vibration causes, and 20 attributes, or vibration phenomena, were used. The system was tested on a centrifugal pump running at 3600 rpm with a misalignment. The overall vibration amplitudes and directions at various locations were given to the user, upon which the user selected which vibration, direction, and location to diagnose. VIBEX returned a list of possible causes, each with a confidence factor. It was left up to the user to decide what the actual cause was based on the information given [33].

Tran et al. proposed a machine condition prognosis based on regression trees and a one-step ahead prediction. The proposed system consisted of 4 steps: data acquisition, data splitting, training–validating, and predicting. The data acquisition stage was the accumulation of historical data and vibration data until faults occur. In the second stage, data was divided into 2 types: training and testing. Step 3 utilized Cao's method and the False Nearest Neighbor Method to create and validate a model, which utilized the CART method to create the regression tree. Finally Step 4 was to use this model and a one-step-ahead prediction to forecast the future value. The error between the predicted value and the actual value in the testing data determined the validity of the prediction. The method was tested on a low methane compressor with minimal error [34].

Sakthivel et al. simulated five faults (bearing, seal, impeller, bearing and impeller together, and cavitation) on a 2 hp monoblock centrifugal pump. Piezoelectric accelerometers were used to measure the vibration signals. Statistical parameters were calculated from the measured vibrations, and then normalized. These parameters were mean, standard deviation, standard error, median, sample variance, kurtosis, skewness, range, minimum, maximum, and sum. Using this information, a decision tree was created using the C4.5 algorithm. C4.5 utilizes information gain (difference in entropy) as a criterion for splitting the data. Each end leaf in the tree defines a fault. Results showed 100 % correct classification [35].

Yang et al. utilized a decision tree for diagnosing rotating machinery. The criterion for the selection process in the decision tree was the information-entropy evaluation function, which was based on information theory. Fourteen causes/classes of vibration, and 20 attributes of vibration were classified in the tree. The classes were grouped into three main groups: mechanical faults, fluid flow, and electrical faults. The attributes consisted of frequency, amplitude, phase, trend, location of vibration, and different components of the frequency. Sohre's cause-result matrix was used to construct the training set for the decision tree. The method was tested on data obtained from other people's researched work, which involved 2 centrifugal pump systems. Results were varied in the strength of the diagnosis for a fault mode [36].

Bevilacqua et al. used the decision tree algorithm CART to classify data from centrifugal pumps from an LPG plant. Data from 143 different centrifugal pumps were classified according to the CART method. Inputs to CART were operative time, plant type, type of seals, fluid type, soot, temperature, kinematic viscosity, pump head, pump capacity, fluid density, and pump power. Potential groups to classify results under were fluid losses, irregular working, losses due to vibrations, mechanical failures, electrical failures, and failures to start [37].

52.5.2 Fuzzy Systems

Perovic et al. proposed a method to detect pump faults using motor current. The fault conditions that were simulated for the experiment were cavitation, blockage, and a damaged impeller. A hydraulic pump fault, such as those listed, would result in less liquid being pumped, leading to a reduction of the load on the motor, less current, an increase in motor shaft speed, and a decrease in slip. Given the right attributes, it was surmised that a unique “fault signature” can be determined to identify properly the different fault conditions. Eight attributes were utilized for this analysis: motor’s per unit slip, current amplitude, amplitude of the shaft speed, and noise around the frequency peak, which was divided into 5 Hz units. All data was scaled to values between 0 and 1 before being inputted into the fuzzy logic system. The fuzzy logic approach was chosen due to the nonlinear nature of the system. The membership functions of the 8 inputs were given, ranging from 3 to 5 regions. This translates into 54 rules for the system, which was then placed through a centroid defuzzification method to generate the output. The output of the fuzzy system was two variables, which describe one of ten different possible conditions, having either one or two faults happening simultaneously. On the APV pump, the rate of success was 92 %, with 37 out of 465 errors classified as “serious identification errors”. On the HMD pump, depending on the case, success rates ranged from 61 % to 92.9 % [38].

Sakthivel et al. studied if it was possible to diagnose a faulty condition in a mono-block centrifugal pump, and if so, to segregate the faults into bearing fault, seal fault, impeller fault, seal and impeller fault together, and cavitation. One accelerometer was used, mounted on the pump inlet. The features extracted from the vibrations were the mean, standard error, median, standard deviation, sample variance, kurtosis, range, minimum, maximum, and sum. C5.0 algorithm was used in order to create the decision tree. The end leaves of the tree designate different faults in the system. The rules to make the membership functions for the fuzzy system were then extracted from the decision tree. This was compared to creating membership functions based on rules created using rough sets. Finally, a PCA based decision tree was created, and the rules were extracted to create the membership functions for a third fuzzy system. The decision tree-fuzzy hybrid system proved to be better than the other two [39].

Azadeh et al. created a fuzzy inference system to diagnose pump failure. The OREDA (Offshore Reliability Data) handbook was used for taxonomy of equipment, as well as quantitative and qualitative information for reliability, availability, safety analysis, and maintenance. In this case, a Mamdani-type fuzzy rule-based inference system was proposed. Pump failure modes were classified into 4 different categories: critical failure, degraded failure, incipient failure, and unknown failure. The inputs to the system were the flow rate, discharge pressure, NPSHR, BHP, efficiency, vibration, and temperature. Triangular membership functions were used for the inputs, and the smallest of maxima method was used for the defuzzification technique for the output. Matlab 7’s fuzzy logic toolbox was utilized to develop the fuzzy inference system. The system was applied to a centrifugal pump of an aromatic plant of a petrochemical complex. For the ten trials given to the system, it was able to correctly diagnose the failure cause, which coincided with those assessed in the field reports [40].

52.6 Combination of Different Methods

Different methods have different strengths that allow them to identify certain fault modes and not others. As a result some researchers have combined various methods in an attempt to identify many fault modes with better accuracy. In most cases, fuzzy logic was coupled with a data driven method in the hybrid system.

Wang et al. created a diagnostic algorithm that utilizes wavelet transforms, rough sets, and partially linearised neural networks (PNN). The ReverseBior wavelet function was used to extract fault features from the measured vibration signals. A 3.7 kW Honda Centrifugal Pump with 7.5 m^3/h capacity was fitted with a vacuum gauge and pressure gauge on the suction and delivery side, as well as six accelerometers. Two sensors were placed at the pump inlet, two at the pump outlet, one on the motor and one on the pump housing. The sensor data was normalized, and then used to create ten non-dimensional symptom parameters (NSPs) in the amplitude domain. The number of NSPs was decreased using rough sets to eliminate redundant information. The remaining NSPs were inputs to the PNN, which was trained using back propagation. When tested, the system was able to correctly detect the normal state with 79.3 % accuracy, cavitation state with 85.1 % accuracy, misalignment state with 14.8 % accuracy, and unbalance state with 61.2 % accuracy [41].

Wang et al. [42], utilized a system similar to that in [41], changing the analysis of the performance of the NSPs. The sensitivity of the NSPs were evaluated using synthetic detection index (SDI), which was also used to determine what type of fault was found in the system.

Wang et al. [43], proposed four methods to aid in the intelligent diagnosis of plant machinery. The first method was a reiteration of their previous paper [41], in which a diagnostic algorithm utilizes wavelet transforms, rough sets, and partially linearized neural networks. The second method utilized a sequential diagnosis method, and was applied to bearing diagnosis for verification. Symptom parameters (SPs) were defined from the signals obtained from the machine, and the ones which were found to be highly sensitive were used for fault diagnosis. The sensitivity equations were given as two equations, labeled distinction index, and distinction rate. The SPs were inputs into a fuzzy neural network. The third method utilized possibility theory, which can be obtained from the probability density function of the machine, and certainty factors. Finally, an adaptive noise cancelling method was applied to the SPs prior to being inputted into a PNN. The results were found to be better than using a high pass filter.

Klema et al. proposed a diagnostic scheme that consisted of signal sensing, signal pre-processing, feature extraction, classification, and the presentation of the diagnosis. Pressure sensors were not used since they were considered intrusive to the system, needing to contact the fluid through tapping the system. As a result, vibrations were acquired using two accelerometers that were attached to the pump casing. Five cases were investigated: normal condition, incipient cavitation, tip vortex cavitation, moderate cavitation, and severe cavitation. Three distinguishing features were extracted from the vibration: power spectral density of the frequency bands, amplitude of the first 22 harmonic components of shaft rotation frequency, and frequency of the rotation of the pump shaft. Classification was performed using four different methods: decision trees, random forest, neural networks, and support vector machines. The inputs to the classifiers were given in one of three ways: an ordinal approach, a standard unordered approach, and the result of using a regression algorithm. The Radviz technique of multidimensional visualization was implemented to provide a visual aid to the results. The method was attempted on a Durco mark III 1K1.5x1–8 pump. Experimental verification of the proposed methodology suggests that it is possible to diagnose cavitation using this type of monitoring system [44].

Zhao et al. proposed a method utilizing fuzzy logic, PCA, and rough sets to diagnose the condition of machines. Features were obtained from the machine, and then evaluated using a fuzzy preference approximation quality (FPAQ) which utilizes rough sets. These features were then sequenced according to their highest value, and placed in groups into a PCA function. The first principle component was first used as a single indicator of the machine health condition. If this component was found not to be a good predictor, the number of features was increased by 1, and then placed into the PCA function again. This method was applied to a centrifugal pump, which utilized 8 features (Amplitude at 1x, 2x, 5x, and 10x; root mean square between 0 and 1x; RMS between 0 and 2X, 0 and 5X, 0 and 10X). These features were extracted from 2 tri-axial accelerometers for both their forward whirl and backward whirl. In total, 144 features were generated and placed as inputs into the system. Using the proposed method, 25 features were selected to generate an indicator [45].

Wolfram et al. used a neuro-fuzzy model based on the Takagi–Sugeno fuzzy system. This system was tested on a centrifugal pump, having model inputs as the inlet and outlet pressures, the flow rate, and the angular speed. The model consisted of three differential equations for the delivery head, motor torque, and torque lost to friction which was all approximated using neuro-fuzzy models. Residuals were calculated based on outputs of the neuro-fuzzy model and those of the actual components, and the model was corrected as a result. Faults such as bearing wear, defect impeller, shaft eccentricities, etc. were seeded into the system, but a unique pattern to identify faults due to residuals and changes in variables in a fault symptom table was not obtained. It was surmised that this difficulty may be a result of the estimated parameters being dependent on the operating point and the applied input signal [46].

Zouari et al. created a real-time fault detection tool for a centrifugal pump using neural and fuzzy techniques, which focused on the following fault modes: partial flow rates, loosening of front/rear pump attachments, misalignment, cavitation, and air injection on the inlet. Initially, features were extracted from machine vibrations with time and spectral analysis. Data analysis methods were then used to reduce the features from 26 to 13 via data analysis methods, such as Fischer, PCA, and FDA. Using the Fisher criterion, the input features were then reduced to 3 parameters, which contained 97 % of the information. The features were then used as inputs to a neuro-fuzzy network, utilizing the neural networks and Fuzzy logic toolboxes in Matlab, and the ANFIS algorithm, created by S. Li and M.A. Elbestawi. The solution to the system was presented with the help of LabVIEW. No testing of the system was performed [47].

52.7 Conclusion

There is still a huge need for an online system in industry that can perform diagnostics and prognostics of centrifugal pumps in real-time. Current industry systems utilize sensors with thresholds that inform the user when possible problems arise in the system. Difficulty occurs in monitoring certain aspects of the system such as lubricants and corroding elements, monitoring vibrations of various components within the system affecting sensors throughout the system, creating a system that can adapt to new problems, and surviving the harsh environment in some industries. As a result, online diagnostic and prognostic systems have been set aside in exchange for more traditional methods of maintenance. Collaboration between research groups and industrial partners is needed in order to pave the way for an automated pump health monitoring system that would be able to interpret sensor and system process data and report a diagnosis and a possible RUL of the system, especially since most pumps lie in remote locations and are found in harsh conditions.

Acknowledgments This research was conducted within the CRC for Infrastructure and Engineering Asset Management, established and supported under the Australian Government's Cooperative Research Centres Programme.

References

1. Karassik I, McGuire JT (1998) Centrifugal Pumps. Chapman & Hall, London
2. Heng A, Zhang S, Tan ACC, Mathew J (2009) Rotating machinery prognostics: state of the art, challenges and opportunities. Mech syst sign process 23(3):724–739
3. Jardine AKS, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech syst sign process 20(7):1483–1510
4. Hashemian HM (2011) Wireless sensors for predictive maintenance of rotating equipment in research reactors. Ann Nucl Energy 38(2–3):665–680
5. Peng Y, Dong M, Zuo M (2010) Current status of machine prognostics in condition-based maintenance: a review. Int J Adv Manufact Technol 50(1):297–313
6. McKee KK, Forbes G, Mazhar I, Entwistle R, Howard I (2011) A review of major centrifugal pump failure modes with application to the water supply and sewerage industries. Proceedings of the ICOMS Asset Management Conference. p 32
7. Beebe RS (2004) Predictive maintenance of pumps using condition monitoring. Elsevier Advanced Technology, Amsterdam
8. Perez RX, Akins RA, Lee CE, Taylor HF (1996) Fiber-optic pressure sensors detect cavitation and flow instabilities in centrifugal pumps. World pumps 359:28–33

9. Mba D, Rao RBKN (2006) Development of acoustic emission technology for condition monitoring and diagnosis of rotating machines: bearings, pumps, gearboxes, engines, and rotating structures. Shock Vibr Dig 38(1):3–16
10. Cudina M (2003) Detectioin of cavitation phenomenon in a centrifugal pump using audible sound. Mech Syst Sign Process 17(6):1335–1347
11. Thobiani AA, Gu F, Ball A (2010) The monitoring of cavitation in centrifugal pumps based on the analysis of vibro-acoustic measurements. Proceedings of the Seventh International Conference on Condition Monitoring and Machinery Failure Prevention Technologies; Stratford-upon-Avon, UK, p 22–24
12. Kallesoe CS, Izaili-Zamanabadi R, Rasmussen H, Cocquempot V (2004) Model based fault diagnosis in a centrifugal pump application using structural analysis. Proc 2004 IEEE Int Conf Control Appl 2:1229–1235
13. Kallesoe CS, Cocquempot V, Izadi-Zamanabadi R (2006) Model based fault detection in a centrifugal pump application. IEEE Trans Control Syst Technol 14(2):204–215
14. Isermann R (1984) Process fault detection based on modeling and estimation methods—A survey. Automatica 20(4):387–404
15. Nordmann R, Aenis M (2004) Fault diagnosis in a centrifugal pump using active magnetic bearings. Int J Rotating Mach 10(3):183–191
16. Cempel C (1988) Vibroacoustical diagnostics of machinery: an outline. Mech Syst Sign Pocess 2(2):135–151
17. Sinha JK, Rao AR (2006) Vibration based diagnosis of a centrifugal pump. Struct Health Monit: An Int J 5(4):325–332
18. Lee S (2002) Cavitation mode analysis of pump inducer. J Mech Sci Technol 16(11):1497
19. Zhang S, Mathew J, Kennedy J, Ma L, Tan A, Anderson D, Hodkiewicz M (2006) Machinery Condition Prognosis Using Multivariate Analysis. In: Mathew J, Kennedy J, Ma L, Tan A, Anderson D (eds) Engineering asset management. Springer, London, pp 847–854
20. Heng A, Tan ACC, Mathew J, Montgomery N, Banjevic D, Jardine AKS (2009) Intelligent condition-based prediction of machinery reliability. Mech Syst Sign Process 23(5):1600–1614
21. Tsai TM, Wang WH (2009) Diagnosis of mechanical pumping system using neural networks and system parameters analysis. J Mech Sci Technol 23(1):124–135
22. Rajakarunakaran S, Venkumar P, Devaraj D, Rao KSP (2008) Artificial neural network approach for fault detection in rotary system. Appl Soft Comput 8(1):740–748
23. Saxena A, Saad A (2007) Evolving an artificial neural network classifier for condition monitoring of rotating mechanical systems. Appl Soft Comput 7(1):441–454
24. Ugechi CI, Ogbonnaya EA, Lilly MT, Ogaji SOT, Probert SD (2009) Condition-based diagnostic approach for predicting the maintenance requirements of machinery. Engineering 1(3):177–187
25. Isermann R (1993) Fault diagnosis of machines via parameter estimation and knowledge processing—tutorial paper. Automatica 29(4):815–835
26. Ebersbach S, Peng Z (2008) Expert system development for vibration analysis in machine condition monitoring. Expert Syst Appl 34(1):291–299
27. Parrondo JL, Velarde S, Santolaria C (1998) Development of a predictive maintenance system for a centrifugal pump. J Qual Maintenance Eng 4(3):198
28. Li J, Khoo L, Tor S (2006) RMINE: a rough set based data mining prototype for the reasoning of incomplete data in condition-based fault diagnosis. J Intell Manuf 17(1):163–176
29. Kollmar D, Hellmann DH (2001) Feature Selection for a Real-World Learning Task. Proceedings of the Second International Machine Learning and Data Mining in Pattern Recognition. Springer, Berlin/Heidelberg, p 157–172
30. Ebersbach S, Peng Z, Yuan C, Yan X (2009) Machine condition monitoring and remaining life prediction using integrated approach advanced tribology. Proceedings of CIST2008 & ITS-IFToMM2008. p 949–956

31. Garcia MA, Clements J, Shelton W (2004) PUMPX: an expert system for pump diagnosis. Proc 34th ASEE/IEEE Front Educ Conf 2:1–4
32. Lihovd E, Tor Ivar J, Sorli A (1995) ROMEX—A Rulebased System for Fault Diagnosis on Rotating Machinery http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.51.1216
33. Yang BS, Lim DS, Tan ACC (2005) VIBEX: an expert system for vibration fault diagnosis of rotating machinery using decision tree and decision table. Expert Syst Appl 28(4):735
34. Tran VT, Yang BS, Oh MS, Tan ACC (2008) Machine condition prognosis based on regression trees and one-step-ahead prediction. Mech Syst Sign Process 22(5):1179–1193
35. Sakthivel NR (2010) Vibration based fault diagnosis of monoblock centrifugal pump using decision tree. Expert Syst Appl 37(6):4040–4049
36. Yang BS, Park CH, Kim HJ (2000) An efficient method of vibration diagnostics for rotating machinery using a decision tree. Int J Rotating Mach 6(1):19–27
37. Bevilacqua M, Braglia M, Montanari R (2003) The classification and regression tree approach to pump failure rate analysis. Reliab Eng & Syst Saf 79(1):59–67
38. Perovic S, Unsworth PJ, Higham EH (2001) Fuzzy logic system to detect pump faults from motor current spectra. Proc Thirty-Sixth Ann Ind Appl Conf 1:274–280
39. Sakthivel NR, Sugumaran V, Nair BB (2010) Comparison of decision tree-fuzzy and rough set-fuzzy methods for fault categorization of mono-block centrifugal pump. Mech Syst Sign Process 24(6):1887–1906
40. Azadeh A, Ebrahimipour V, Bavar P (2010) A fuzzy inference system for pump failure diagnosis to improve maintenance process: the case of a petrochemical industry. Expert Syst Appl 37(1):627–639
41. Wang H (2009) Intelligent diagnosis method for a centrifugal pump using features of vibration signals. Neural Comput Appl 18(4):397–405
42. Wang H, Chen P (2007) Sequential condition diagnosis for centrifugal pump system using fuzzy neural network—Letters and Reviews. Neural Inf Process 11(3):41–50
43. Wang H (2010) Intelligent diagnosis methods for plant machinery. Front Mech Eng China 5(1):118–124
44. Klema J, Flek O, Kout J, Novakova L (2005) Intelligent Diagnosis and Learning in Centrifugal Pumps. Emerg Solutions Future Manufact Syst 159:513–522 Springer
45. Zhao X (2010) Application of fuzzy preference based rough set model to condition monitoring rough sets and current trends in computing. Springer-Verlag, Berlin
46. Wolfram A, Fussel D, Brune T, Isermann R (2001) Component-based multi-model approach for fault detection and diagnosis of a centrifugal pump. Proc 2001 Am Control Conf 6:4443–4448
47. Zouari R, Sieg-Zieba S, Sidahmed M (2004) Fault detection system for centrifugal pumps using neural networks and neuro-fuzzy techniques. Surveillance 5 CETIM Senlis, p 11–13

Chapter 53
Condition-Based Monitoring of a Centrifugal Pump Using Mahalanobis-Taguchi System

S. Jagannathan, C. Saygin and M. Zawodniok

Abstract In this paper, a condition-based monitoring (CBM) model that uses the Mahalanobis-Taguchi System (MTS) for fault detection, isolation, and prognostics is presented. The proposed scheme fuses data from multiple sensors into a single system level performance metric using Mahalanobis Distance (MD) and generates fault clusters based on MD values. MD thresholds derived from the clustering analysis are used for fault detection and isolation. When a fault is detected, the prognostics scheme, which monitors the progression of the MD values over time, is initiated. Then, using a linear approximation, time to failure is estimated. The performance of the scheme has been validated via experiments performed on a mono-block centrifugal water pump testbed. The pump has been instrumented with vibration, pressure, temperature and flow sensors and experiments involving healthy and various types of faulty operating conditions have been performed. The experiments show that the proposed approach renders satisfactory results for centrifugal water pump fault detection, isolation and prognostics.

S. Jagannathan (✉) · M. Zawodniok
Department of Computer and Electrical Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA
e-mail: sarangap@mst.edu

M. Zawodniok
e-mail: mjzx9c@mst.edu

C. Saygin
Mechanical Engineering Department, University of Texas San Antonio, San Antonio, TX 78249, USA
e-mail: can.saygin@utsa.edu

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_53,

53.1 Introduction

Centrifugal pumps are extensively used in a wide variety of industries [1]. Unexpected failure of centrifugal pumps can become costly due to the interruption in processes and further costs are incurred in terms of repair/replacement. In a centrifugal pump, bearing, seal and impeller are critical components that directly affect the desired performance of the pump. Bearing, seal and impeller defects, and cavitation can result in problems such as abnormal noise, leakage, drop in hydraulic performance, structural vibration and damage on the pump components by corrosion and pitting.

In the literature, impeller [1–7] and seal [5–7] failures are found to be the predominant failure modes. In general, researchers have approached the centrifugal pump performance monitoring and fault diagnosis problem in two ways: (1) utilizing qualitative or quantitative model based techniques, (2) using process history based models. Acoustic emission [8], vibration analysis [1, 9, 10], frequency analysis [4–7, 9–11], and band-width analysis [1, 3] are among the commonly used techniques.

Mahalanobis-Taguchi System (MTS) has been widely used in various diagnostic applications that deal with data classification. In MTS, Mahalanobis Distance (MD) is used to identify the degree of abnormality of data, whereas the Taguchi methods, which utilize orthogonal arrays (OA) and signal-to-noise ratio (S/N), are used to evaluate the accuracy of predictions and to optimize the system [12]. MD, introduced by Mahalanobis in 1936 [13], is a multivariate generalized measure used to determine the distance of a data point to the mean of a group. MD is measured in terms of the standard deviations from the mean of the samples and provides a statistical measure of how well the unknown data set matches with the ideal one. The advantage of the MD is that it is sensitive to the intervariable changes in the reference data. Therefore, it has traditionally been used to classify observations into different groups and diagnosis [14]. A detailed review and critique of the MTS method is presented by Woodall et al., along with a discussion and response section by Jugulum et al. [15].

MTS was developed by Taguchi as a diagnostic technique using multivariate data. The MD introduced by Mahalanobis in 1936 takes the correlation structure of a system into account [16]. Therefore, MTS-based clustering (i.e., classification) is not a new technique. However, the main contribution of this study is the combination of MTS-based clustering with prognostics and its application to pump failures, which has not been studied in the literature. By combining these methods, the benefits of MTS have been expanded over to the prognostics field. More specifically, this paper (1) uses MTS in order to fuse multi-sensor information into a single system performance metric; (2) builds upon MTS by introducing the use of MD-based fault clusters for fault isolation; (3) utilizes the progression of MD values over time in order to estimate the remaining useful life of components (i.e., time to failure); and (4) applies this proof-of-concept model to the domain of centrifugal pump monitoring, diagnostics, and prognostics area.

53.2 Notations

x_{ij}	ith characteristic in the jth observation
n	Number of observations
s_i	Standard deviation of the ith characteristic
Z_{ij}	Normalized value of the ith characteristic in the jth observation
s_z	Standard deviation of the normalized values
C	Correlation matrix
C^{-1}	Inverse of the correlation matrix
MD_j	Mahalanobis Distance for the jth observation
k	Number of characteristics
η_q	Signal-to-noise ratio for the qth run of the orthogonal array (OA)
t	Number of abnormalities under consideration
σ_X	Standard deviation of the random variable X
σ_Y	Standard deviation of the random variable Y
μX	Expected value for X
μY	Expected value for Y
E	Expected value operator
cov	Covariance

53.3 Methodology

In this paper, the proposed model utilizes a Mahalanobis Distance based fault clustering method in order to classify faults into different categories. Using these clusters, fault detection and root cause analysis is performed. The scheme also monitors the progression of the MD values over time in order to facilitate prognosis of time to failure for components.

MTS starts with data collection on normal observations. Then, Mahalanobis Distance is calculated using certain characteristics to investigate whether MD has the ability to differentiate the normal group from an abnormal group. If MD cannot detect the normal group using those particular characteristics, then a new combination of characteristics need to be explored. When the right set of characteristics are found, Taguchi methods are employed to evaluate the contribution of each characteristic. If possible, dimensionality is reduced by eliminating those characteristics that do not add value to the analysis. MTS consists of four stages [12–14]:

53.3.1 Stage 1: Construction of Mahalanobis Space

The first step in MTS is the construction of the Mahalanobis Space (MS) as the reference. In order to construct MS, the variables that represent “normal” conditions are defined and sampled. In this study, new bearings that are properly lubricated represent the normal case. The steps for the construction of MS are outlined below:

1. Calculate the mean for each characteristic in the normal data set as:

$$\bar{x}_i = \frac{\sum_{j=1}^{n} X_{ij}}{n} \tag{53.1}$$

2. Then, calculate the standard deviation for each characteristic:

$$s_i = \sqrt{\frac{\sum_{j=1}^{n} \left(X_{ij} - \bar{x}_i\right)^2}{n-1}} \tag{53.2}$$

3. Next, normalize each characteristic, form the normalized data matrix $\left(z_{ij}\right)$ and take its transpose $\left(Z_{ij}^T\right)$

$$Z_{ij} = \frac{\left(X_{ij} - \bar{x}_i\right)}{s_i} \tag{53.3}$$

4. Then, verify that the mean of the normalized data is zero:

$$\bar{z}_i = \frac{\sum_{j=1}^{n} Z_{ij}}{n} = 0 \tag{53.4}$$

5. Verify that the standard deviation of the normalized data is one:

$$s_z = \sqrt{\frac{\sum_{j=1}^{n} \left(Z_{ij} - \bar{z}_i\right)^2}{n-1}} = 1 \tag{53.5}$$

6. Form the correlation matrix C for normalized data. Calculate the matrix elements C_{ij} as follows:

$$C_{ij} = \frac{\sum_{m=1}^{n} \left(Z_{im} Z_{jm}\right)}{n-1} \tag{53.6}$$

7. Calculate the inverse of the correlation matrix (C^{-1}).
8. Finally, calculate MD as:

$$MD_j = \frac{1}{k} Z_{ij}^T C^{-1} Z_{ij} \tag{53.7}$$

53.3.2 Stage 2: Validation of MS

In order to validate MS, observations that fall outside the normal group are identified and the corresponding MD values are calculated. The characteristics of the abnormal group are normalized using the mean and the standard deviation of the corresponding characteristics in the normal group. The correlation matrix corresponding to the normal group is used to compute the MDs of the abnormal cases. If MS has been constructed using the appropriate characteristics, then the MDs of the abnormal group should have higher values than that of the normal group.

53.3.3 Stage 3: Identification of the Useful Characteristics

In this stage, Orthogonal Arrays (OA) and Signal-to-Noise (SN) ratios are used to select the critical variables. OA is used to estimate the effects of features and the effect of interactions by minimizing the number of experiments. In MTS, each variable is assigned to one column and is set with two levels, that is, level 1 is corresponding to the presence of a variable, and level 2 is corresponding to the absence of a variable. In other words, the second phase is Taguchi's robust engineering, which is applied to determine the important variables and then optimize the system.

An orthogonal array is a table that lists the combination of characteristics, which enables testing the effects of the presence or absence of a characteristic. The size of the OA is determined by the number of characteristics and the levels they can take. For the abnormal cases, MD values are calculated using the combination of the characteristics dictated by the OA and the larger-the-better signal-to-noise ratio is calculated as follows [12, 14]:

$$\eta_q = -10\log\left[\frac{1}{t}\sum_{j=1}^{t}\frac{1}{MD_j}\right] \tag{53.8}$$

Then, an average signal-to-noise ratio at level-1 and level-2 of each characteristic is obtained and the gain in signal-to-noise ratio values for each characteristic is calculated [16]:

$$Gain = (Avg.\,S/N\,Ratio)_{Level-1} - (Avg.\,S/N\,Ratio)_{Level-2} \tag{53.9}$$

The characteristics with positive gain are selected for the detection of abnormalities and the rest are discarded.

53.3.4 Stage 4: Decision Making

In the final stage, the application under investigation is monitored via data collection using the MS. MDs are calculated and if MD $\gg 1$, it is concluded that the application is displaying abnormal behavior and appropriate corrective actions need to be taken. If MD ≤ 1, then the conditions are normal.

53.4 Condition-based Monitoring

A fault is a change from the normal condition of a system. Fault detection is the first step in diagnostics, which indicates the occurrence of a fault in a monitored system. In this study, an MD-based data clustering technique is used in order to classify centrifugal pump failure data into different fault groups. The technique takes advantage of the fact that the Mahalanobis Distance is calculated using the inverse of the correlation matrix (Eq. 53.6). The correlation matrix consists of the correlation coefficients, which is a normalized measure of the strength of the linear relationship between variables as shown below:

$$r_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X-\mu X)(Y-\mu Y))}{\sigma_x \sigma_Y} \tag{53.10}$$

The correlation between the MDs can be visualized by plotting the current MD value against the previous MD value. In this case, data from the same failure are clustered together due to their correlation, as shown in Fig. 53.1.

Upon completion of the clustering analysis, the appropriate thresholds are selected based on a variety of techniques, such as the mean and three times the standard deviation of the MD values for each fault cluster, probabilistic thresholding method [16], or bisection method [17]. Determining a threshold is an important step in carrying out a diagnosis and prognosis process effectively. This paper uses the mean and $\pm$ three times the standard deviation approach to set thresholds.

Prognostics phase of the study involves (a) detection of an abnormal condition or fault, (b) identifying the root cause or fault isolation, and, (c) estimating the remaining useful life or time to failure. In the prognostics case, the MD values are calculated using a predetermined time window as the process continues. Fault

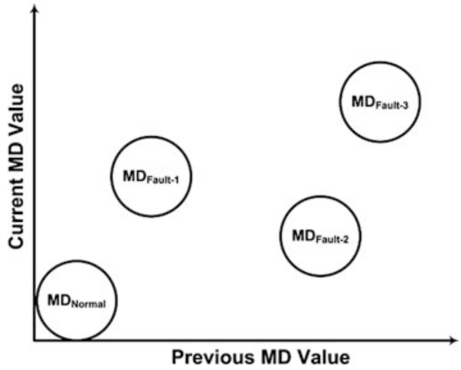


Fig. 53.1 MD-based fault clustering

detection occurs once the MDs leave the normal regime and the tracking of the MD trend is initiated. By observing the direction of the transition of the MD trend between the normal case and one of the fault clusters, the possible root cause is identified. This is accomplished by calculating the angle θ between the unknown point U and the mean MD value of each of the known fault clusters assuming the fault progression will follow a roughly linear trajectory, as shown in Fig. 53.2.

Once the MD data enters the root cause failure cluster, prognosis of the time to failure is initiated via linear approximation. First, the slope of the progression of the MD value is generated using the MD value of the current and the previous time windows. Then, using the slope and the current MD value, the time to failure is determined at each instant of time (53.11). The failure threshold is selected by the designer when the performance of the machine is unsatisfactory.

$$slope = \frac{MD_t - MD_{failure-threshold}}{time_t - time_{failure}} \tag{53.11}$$

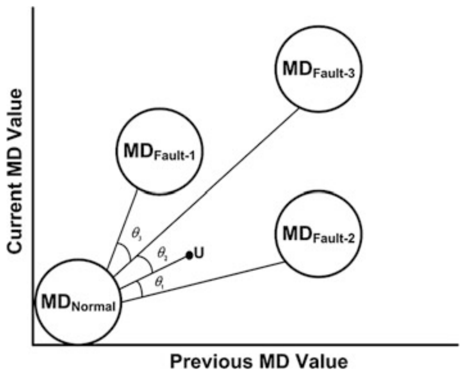


Fig. 53.2 Illustration of angle-based fault isolation

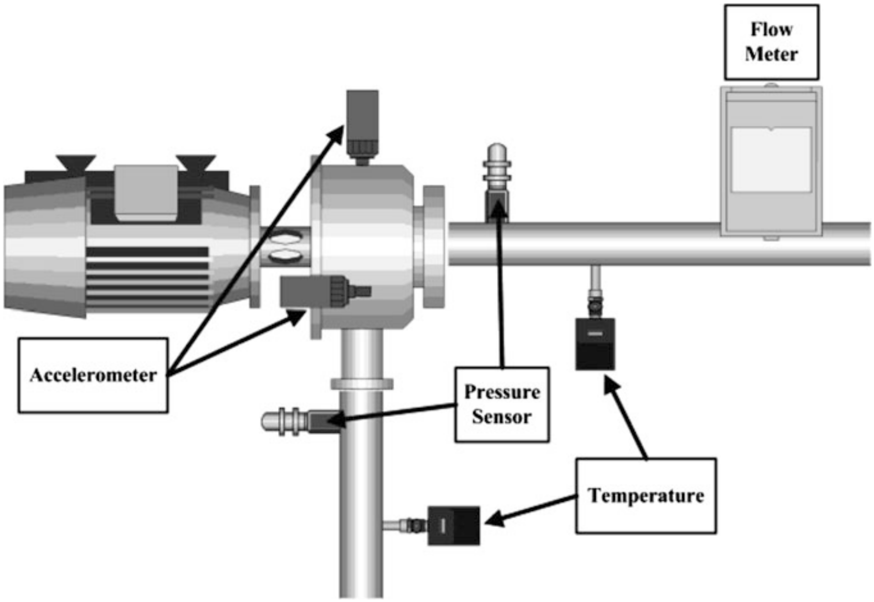


Fig. 53.3 Centrifugal pump setup

53.5 Experimentation

The centrifugal pump setup is shown in Fig. 53.3. The 1/2 HP centrifugal pump has been instrumented with pressure and temperature sensors at the inlet and pressure, temperature, and flow sensors at the outlet. In addition, two accelerometers are installed in the pump casing in order to measure the lateral and vertical vibration.

The water pump testbed was operated for 150 h in order to collect the signature for normal operating conditions. After the initial data collection for normal conditions was complete, three types of pump failures one at a time were seeded in order to accelerate the testing. The pump failures include damaged seal, eroded impeller, and clogged filter.

The raw sensor measurements are used as the input parameters for the MTS model, including lateral acceleration, vertical acceleration, inlet pressure, outlet pressure, and outlet flow. The $L_8(2^7)$ orthogonal array is utilized for the S/N analysis. Using the larger-the-better type signal to noise ratios and the overall gains, all of the parameters mentioned above are selected as the final input parameters for further analysis since none of the inputs could be eliminated in order to be able to detect all three failures. Figure 53.4 shows the results for the OA analysis for the experiments.

Using the mean and ± three times the standard deviation of the MDs for each fault type, the thresholds for each fault cluster are identified:

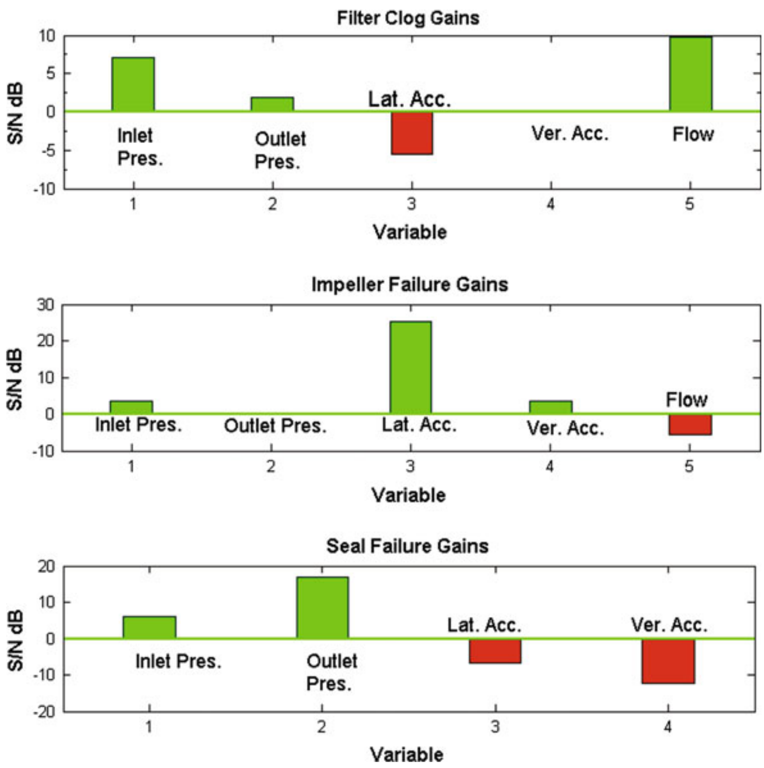


Fig. 53.4 Orthogonal array (*OA*) analysis gains

- Normal Condition: MD_min = 0.02, MD_max = 10, StdDev_MD = 0.89
- Filter Clog Failure: MD_min = 482, MD_max = 582, StdDev_MD = 38
- Seal Failure: MD_min = 561, MD_max = 827, StdDev_MD = 59
- Impeller Failure: MD_min = 28,430, MD_max = 29,179, StdDev_MD = 155

There is an overlap on MD values for filter clog and seal failures (561–582), which is not an ideal case for an MTS application. The scheme will still be able to detect an abnormal condition for an MD value that fall in this range; however, the failure type cannot be determined specifically. In order to implement the MTS model for condition-based monitoring and determine the direction of actual progression of a fault based on the fault clusters, transition data is utilized following the scheme shown in Fig. 53.5.

The proposed scheme exploits the idea of using each of the fault cases as a reference in the generation of fault clusters. At time = 1, test data is acquired and the MD values are calculated using the normal case as the basis for comparison. Fault detection occurs when the MD of the current process crosses the predetermined threshold of the normal case. If a fault is detected, a flag is set to initiate the parallel calculation of MD values with the faulty cases as the references, and

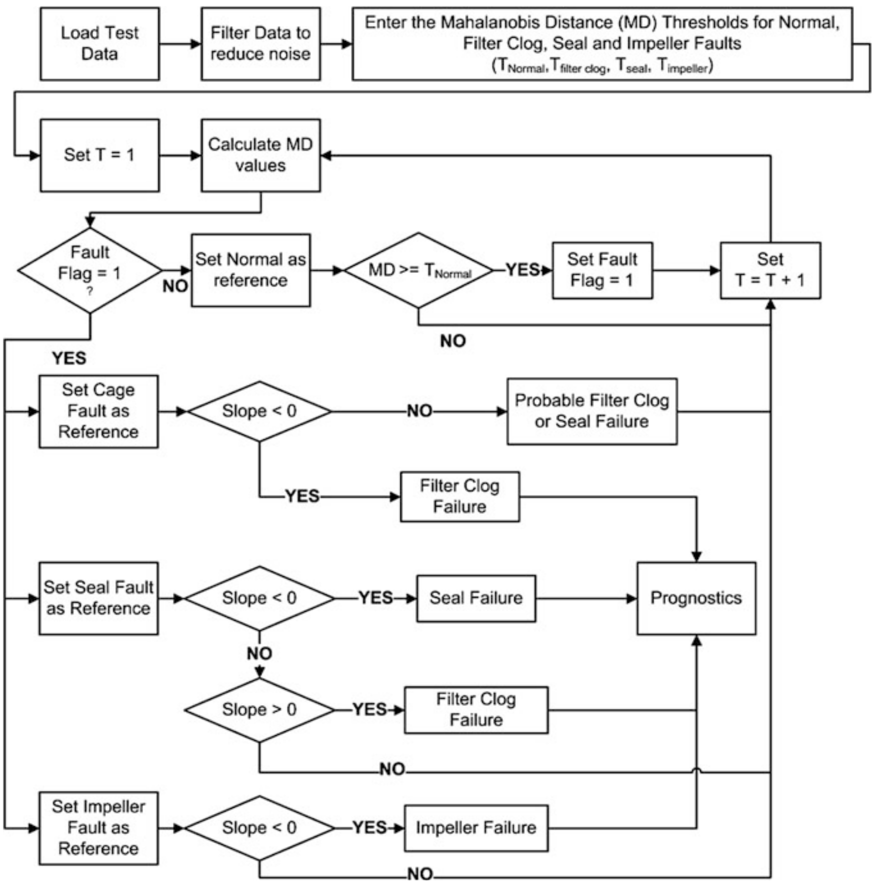


Fig. 53.5 Condition-based monitoring scheme

the time is advanced to the next time window. Then, by using the four references, the direction of progression of the fault to its fault cluster is determined and the root cause identified.

Figure 53.6 shows the time evolution of MD values for the seal failure experiment. Figure 53.7 shows the time-to-failure (TTF) estimation plot for the same experiment using a 30 min time window using Eq. (53.11). The failure threshold value selected for the experiment is 827. The estimation is triggered as soon as a particular operation enters a fault cluster and stays within the cluster for a period of two time windows in order to minimize false fault classification alarms. Figure 53.8 shows the time evolution of MD values for the filter clog failure experiment, while Fig. 53.9 shows the time to failure estimation for the same experiment using a 30 min time window and a failure threshold value of 582.

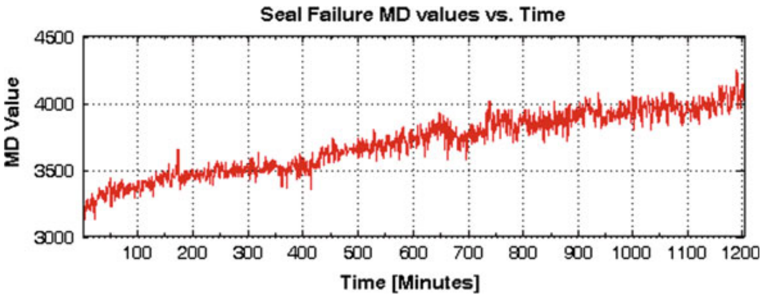


Fig. 53.6 Time evolution of MD values for the seal failure experiment

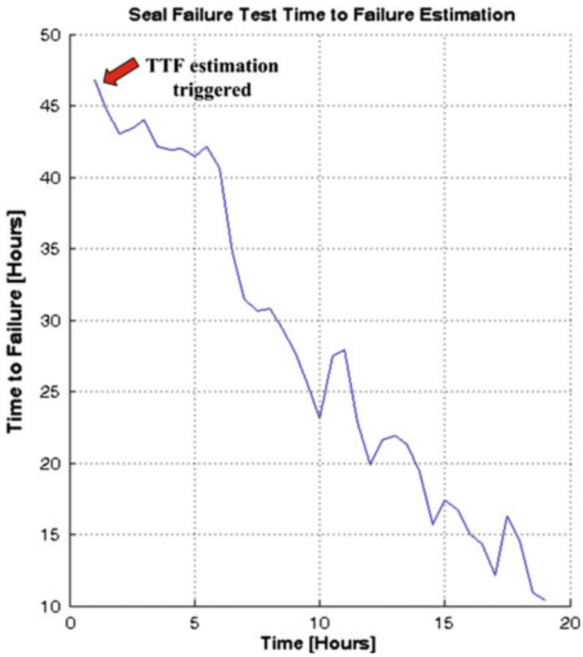


Fig. 53.7 Time to failure estimation for the seal failure experiment

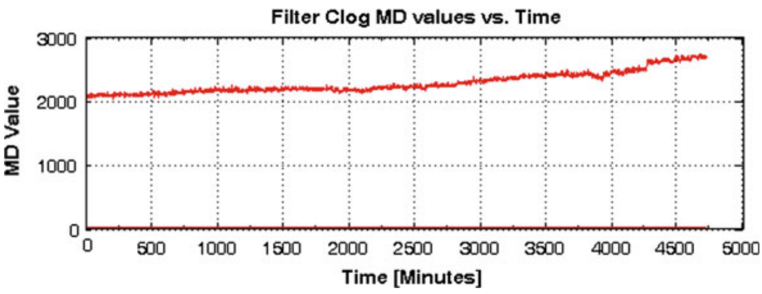
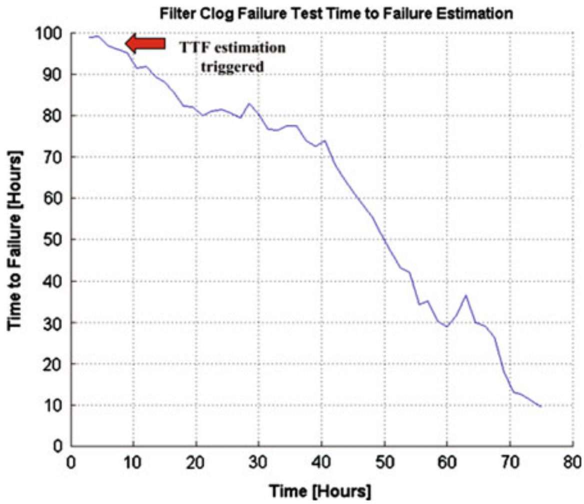


Fig. 53.8 Time evolution of MD values for the filter clog failure experiment

Fig. 53.9 Time to failure estimation for the filter clog failure experiment



53.6 Conclusions

In this paper, a condition-based monitoring scheme founded on Mahalanobis-Taguchi System is presented. The proposed scheme utilizes Mahalanobis Distance for fault clustering and the progression of MD values for diagnostics and prognostics. The performance of the scheme has been demonstrated through experiments performed on a centrifugal pump setup. Advantages of the proposed approach can be summarized as follows:

- The proposed approach fuses all of the pertinent information obtained from a multi-sensor environment into a single performance metric, which is the Mahalanobis Distance.
- The proposed approach can function equally well in both the data and feature domains. In the case when the raw data collected from sensors is adequate to differentiate between various system health conditions, the data can directly be used as inputs to the system. If the raw data is not sufficient, then higher level features can be extracted from the data by further analysis, and these features can be utilized as inputs.
- It provides a unique solution for centrifugal pump fault detection, isolation and prognostics, eliminating the need for developing a tool for each one separately.
- It provides a methodical way to identify the features critical for the process, reducing dimensionality of the problem. Therefore, the analysis overhead is reduced since efforts can be concentrated on the key features.
- Since the proposed scheme is not computationally exhaustive, it can easily be implemented onto wireless motes and deployed to perform online monitoring (i.e., real-time) and decision making in a variety of industrial environments.

The analysis and results presented in this paper, especially the thresholds, are valid for the particular experimental setup and operating conditions. However, the proposed approach can be applied to other processes.

References

1. Lihovd E, Johannessen TI, Steinebach C, Rasmussen M (1998) Intelligent diagnosis and maintenance management. J Intell Manuf 9(6):523–537
2. Byskov RK, Jacobsen CB, Pedersen N (2003) Flow in a centrifugal pump impeller at design and off-design conditions—Part II: large eddy simulations. J Fluids Eng 125(1):73–83
3. Byington CS, Watson M, Edwards D, Dunkin B (2003) In-line health monitoring system for hydraulic pumps and motors. In: Proceedings of the IEEE aerospace conference 2003, vol 7, pp 3279–3287
4. Koo IS, Kim WW (2000) The development of reactor coolant pump vibration monitoring and a diagnostic system in the nuclear power plant. ISA Trans 39(3):309–316
5. Perovic S, Unsworth PJ, Higham EH (2001) Fuzzy logic system to detect pump faults from motor current spectra. Industry applications conference—Thirty-Sixth IAS annual meeting, vol 1, pp 274–280
6. Sakthivel NR, Sugumaran V, Babudevasenapati S (2010) Vibration based fault diagnosis of monoblock centrifugal pump using decision tree. Expert Syst Appl 37(6):4040–4049
7. Sakthivel NR, Sugumaran V, Nair BB (2010) Comparison of decision tree-fuzzy and rough set-fuzzy methods for fault categorization of mono-block centrifugal pump. Mech Syst Signal Process 24(6):1187–1906
8. Langthjem MA, Olhoff N (2004) A numerical study of flow-induced noise in a two-dimensional centrifugal pump: Part I hydrodynamics. J Fluids Struct 19(3):349–368
9. Dong R, Chu S, Katz J (1997) Effect of modification to tongue and impeller geometry on unsteady flow, pressure fluctuations, and noise in a centrifugal pump. J Turbomach 119(3):506–515
10. Lebold MS, Maynard K, Reichard K, Trethewey M, Hasker J, Lissenden C, Dobbins D (2005) A non-intrusive technique for on-line shaft crack detection and tracking. In: Proceedings of the IEEE aerospace conference, p 1–11
11. Zhang S, Asakura T, Xu X, Xu B (2003) Fault diagnosis system for rotary machine based on fuzzy neural networks. Japan Soc Mech Eng 46(3):1035–1041
12. Taguchi G, Chowdury S, Wu Y (2001) The Mahalanobis Taguchi System. McGraw Hill, New York
13. Mahalanobis PC (1936) On the generalized distance in statistics. In: Proceedings, National Institute of Science of India, vol 2(1), pp 49–55
14. Taguchi G, Jugulum R (2002) The Mahalanobis-Taguchi strategy—A pattern technology system. Wiley, New York
15. Woodall WH, Koudelik R, Tsui KL, Kim SB, Stoumbos ZG, Carvounis CP, Jugulum R, Taguchi G, Taguchi S (2003) A review and analysis of the Mahalanobis-Taguchi System. Technometrics 45(1):1–15
16. Su CT, Hsiao YH (2007) An evaluation of the robustness of MTS for imbalanced data. IEEE Trans Knowl Data Eng 19(10):1321–1332
17. Yang T, Cheng YT (2010) The use of Mahalanobis-Taguchi System to improve flip-chip bumping height inspection efficiency. Microelectron Reliab 50(3):407–414

Chapter 54
Experimental Validation of LS-SVM Based Fault Identification in Analog Circuits Using Frequency Features

A. S. S. Vasan, B. Long and M. Pecht

Abstract Analog circuits have been widely used in diverse fields such as avionics, telecommunications, healthcare, and more. Detection and identification of soft faults in analog circuits subjected to component variation within standard tolerance range is critical for the development of reliable electronic systems, and thus forms the primary focus of this paper. In this paper, we have experimentally demonstrated a reliable and accurate (99 %) fault diagnostic framework consisting of a sweep signal generator, spectral estimator and a least squares-support vector machine. The proposed method is completely automated and can be extended for testing other analog circuits whose performances are mainly determined by their frequency characteristics.

54.1 Introduction

A survey conducted in the area of diagnostic testing of electronic circuits and devices indicated that 80 % of the faults occur in the analog segment [1, 2]. Existing diagnostic testing methods for analog circuits are largely dependent on simulation data and their accuracy is affected by component tolerances, the complex nature of the fault mechanisms, nonlinearity issues and the deviation of

A. S. S. Vasan · B. Long · M. Pecht (✉)
Center for Advanced Life Cycle Engineering (CALCE), University of Maryland, College Park 20742, USA
e-mail: pecht@calce.umd.edu

B. Long
School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, China

M. Pecht
Center for Prognostics and System Health Management, City University of Hong Kong, Kowloon, Hong Kong

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_54,

component parameters due to operational and environmental stresses [3]. Detecting and isolating soft faults when components vary within their tolerance range is a challenging fault diagnosis problem [4]. Situations in which the circuit topology is not affected but deviation in circuit elements from their nominal value leads to changes in the operational range of the circuit thereby affecting its performance are referred to as *soft faults* [3]. On the other hand, *component tolerance* does not result in a fault but can affect performance of diagnostic techniques through small changes in circuit's component parameters within their standard tolerance range [5]. Over the past 10 years, various diagnostic methods have been proposed. These are either based on derived diagnostic equations at selected nodes of the circuit [5–9] or on simulated fault data (data-driven) [10–13]. Fault diagnosis based on the former approach suffers from drawbacks associated with the estimation procedure when dealing with nonlinear circuits and are limited by the poor accessibility of internal nodes of modern integrated circuits.

Neural networks (NNs) are the most popular data-driven methods used in soft fault diagnosis in analog circuits. The advantage of NNs is that they do not need a fault model, which is hard to obtain for analog circuits. However, a NN is hard to control and requires sufficiently large amount of fault data for training. It also has other disadvantages such as low convergence rate, local optimal solution, and poor generalization when the number of fault samples is limited [13–15]. Feeding the extracted features directly as the input to a NN can result in a large NN, even for a small circuit [10]. Improved performance with NNs has been reported [11, 12] at the cost of additional computations for data preprocessing. In recent years, researchers have demonstrated SVM as an effective tool for diagnostics of analog circuits [9, 14, 15, 19, 20]. Support vector machine (SVM) is a machine learning tool that accounts for the tradeoff between learning ability and generalizing ability by minimizing structure risk [16–18]. Most of the aforementioned methods [5–13] are based on simulations data which are not consistent with real data [11, 12]. Also, they are validated using data from faulty circuits, where the faulty components vary from their nominal value by 50 %. The diagnosability of these approaches under component variations just outside tolerance range is still questionable. In our research, we experimentally demonstrate 99 % diagnosability using the least squares support vector machine (LS-SVM) even in the unexplored component variation levels. Conventional frequency domain and wavelet features are extracted from the spectral estimate of the circuit's response to a sweep signal, to demonstrate reliable and accurate fault diagnosis with reduced test time. The material in this work is arranged as follows. Section 54.2 gives a brief description of our diagnostic framework followed by a discussion on spectrum estimation and feature selection in Sects. 54.3 and 54.4 respectively. In Sect. 54.5, the multiclass LS-SVM for fault classification is discussed. This is followed by a detailed description on our experiments demonstrating the strength of our diagnostic technique in Sect. 54.6. Finally, conclusions are drawn in Sect. 54.7.

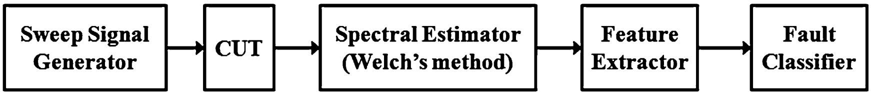


Fig. 54.1 Overview of the proposed diagnostic setup

54.2 Description of the Fault Diagnostic Framework

Figure 54.1 shows the proposed diagnostic system for a circuit under test (CUT). It uses a sweep signal generator as the test signal generator followed by a spectral estimator and a feature extractor and finally a fault classifier based on a multiclass LS-SVM. For most of the electronic circuits, the behavioral characteristics are accompanied by a unique frequency response. In this approach, this feature is exploited for generating fault signatures. In order to obtain the frequency domain features, the CUT is excited by a sweep signal containing a frequency bandwidth which is larger than that of the CUT. Then the associated power spectral density (PSD) of the CUT response is estimated from which frequency domain features are extracted. It is practically not feasible to obtain the true PSD from the observation and only an estimate of the PSD can be obtained. Hence, we employ a nonparametric spectrum estimator-based on Welch's method [22] to estimate the PSD. Then, two types of features, namely (1) the conventional frequency features and (2) the wavelet features are extracted from the estimated PSD and are used as signatures for the fault classifier. In this work, we have used a one-against-one multiclass least squares-support vector machine (LS-SVM) for fault classification.

Analog circuit fault diagnosis procedure involves two phases: the training and the diagnostic phase. The most frequently occurring faults are investigated to define faults of interest. CUT is then simulated for these hypothesized faulty conditions and excited by the test stimuli so as to develop signatures representing fault conditions. The signatures of these responses are then stored in a dictionary for use during the online identification of faults. During the diagnostic phase, the CUT is excited by the test stimuli and the signatures are acquired. These signatures are compared with those stored in the fault dictionary for identification. Generally, the fault classifiers are used in organizing the signatures using a specific criterion to identify the faulty CUT to one of the prestored faults.

54.3 Spectrum Estimation

Let $\mathbf{x}(n); n = 0, \ldots, N-1$ be the observed response of the CUT. Take segments $\mathbf{x}_i(n)$ of length L with starting points of these segments D units apart. If K sequences cover the entire N data points, then $(K-1)D + L = N$. Then a data window $w(n)$ is applied to each of these sequences, thereby producing a set of K modified periodograms that are averaged. The modified periodograms obtained by applying the window is given by

$$\mathbf{I}_k^i\left(e^{j\omega}\right) = \frac{L}{U}\left|\frac{1}{L}\sum_{n=0}^{L-1}\mathbf{x}_i(n)w(n)e^{-jn\omega}\right|^2;\ k = 1, 2, \ldots, K,$$
$$\text{where}\quad \mathrm{U} = \frac{1}{\mathrm{L}}\sum_{\mathrm{n}=0}^{\mathrm{L}-1}|\mathrm{w(n)}|^2. \tag{54.1}$$

Finally the average of these periodograms gives the expected value of the Welch's spectral estimate [22].

54.4 Feature Selection and Extraction

In this proposed diagnostic approach, two types of features are extracted from the PSD of the CUT response to sweep test signal. Both of these features are explained with reference to the well know frequency responses of analog filter circuits. In general there are four types of filters that are used to process analog signals. These are the low pass (LP), high pass (HP), band pass (BP), and band stop (BS) filters. Among these four types, the method to select the frequency features for LP filters can be used for HP filters, and the frequency features for BP filters can be used for BS filters.

54.4.1 Conventional Frequency Features

The conventional frequency features refers to the characteristics components which can identify the circuits behavior in the frequency domain. For a BP filter, the operational characteristics are defined using (1) the center frequency (f_0); (2) 3 dB upper and lower pass-band limits (f_l and f_u); and (3) the maximum of the frequency response $(\mathbf{H}(f_0))$. Similarly, for the LP filters, both the center frequency and the maximum of the frequency response $(\mathbf{H}(f_0))$ become a part of the feature vector. Instead of the 3 dB pass-band limit, the 3 dB cutoff frequency (f_{cut}) is used as the third element of the feature vector. Additional features can also improve the classification accuracy. Hence, the frequency responses at (1) the 3 dB pass-band limit $(\mathbf{H}(f_u))$ for BP filters and (2) the 3 dB cutoff frequency $(\mathbf{H}(f_{cut}))$ for an LP filter are used as additional feature elements. Thus, the conventional frequency features for the BP and LP filters are expressed as:

$$\mathbf{BPF} = [f_0, f_l, f_u, \mathbf{H}(f_0), \mathbf{H}(f_u)] \text{ and } \mathbf{LPF} = [f_0, f_{cut}, \mathbf{H}(f_0), \mathbf{H}(f_{cut})]. \tag{54.2}$$

Similarly, for the HP and BS filters the features are chosen as they are in LP and BP filters, except that the upper cutoff frequency is replaced by a lower cutoff frequency for HP filters and the 3 dB pass-band limit is replaced with a 3 dB stop band limits in BS filters. Also, the minimum frequency response is used in place of the maximum frequency response for BS filters.

54.4.2 Wavelet Features

To improve the classification accuracy, additional features are extracted using wavelet transformation where the frequency response, $P(f)$ is decomposed into the so-called approximation and detail signals using the multiresolution representation [22]

$$P(f) = \sum_{j=1}^{J} W_k^j \psi^j\left(f - 2^j k\right) + \sum_k V_k^J \phi^J\left(f - 2^J k\right). \tag{54.3}$$

where, $\mathbf{y}^j(f)$ is the family of wavelet functions and $\mathbf{f}^J(f)$ is the family of scaling functions. The approximation and detail signals are expressed using the approximation (W_k^j) and detail (V_k^j) wavelet coefficients which are obtained through low and high pass filtering respectively. Based on earlier research [19], we have selected the Haar wavelet and the coefficients of the detail signals of levels 1 through 5 for extracting two types of wavelet features. In the first type, the energy contained in the detail signal at various levels of decomposition is used as a wavelet feature. This is expressed as follows:

$$E_j = \sum_k \left|\mathrm{W}_k^j\right|^2, j = 1, 2, \ldots, J, \tag{54.4}$$

where, W_k^j is the kth coefficient of the detail at j th level of decomposition and J is the predefined decomposition level. The energy indicators are easy to implement, but they fail to utilize complete information from wavelet transformation. Hence, in the second type, we use the means $\left(\mu_j\right)$ and standard deviations $\left(\sigma_j\right)$ at each decomposition level as wavelet features:

$$\mu_j = \frac{1}{n}\sum_{k=1}^{n} \mathrm{W}_k^j \text{ and } \sigma_j = \sqrt{\mathrm{E}\left\{\mathrm{W}_k^j - \mu_j\right\}}, \tag{54.5}$$

where, n denotes the predefined number of coefficients used as features. Thus, the final feature vector including the features extracted using wavelet decomposition for a BP filter looks like either

$$\mathbf{BPF} = [f_0, f_l, f_u, \mathbf{H}(f_0), \mathbf{H}(f_u), E_1, \ldots, E_5], \tag{54.6}$$

or

$$\mathbf{BPF} = [f_0, f_l, f_u, \mathbf{H}(f_0), \mathbf{H}(f_u), \mu_1, \ldots, \mu_5, \sigma_1, \ldots, \sigma_5]. \tag{54.7}$$

In this work, we have evaluated different test conditions which are based on (1) the frequency feature type, (2) the wavelet feature type, and (3) the normalization method. By frequency feature type we mean that the feature vector constitutes either the conventional frequency features of the filter circuit, or the wavelet features, or a combination of both. When wavelet features are chosen, either the energy indicators, or the mean and standard deviation of the detail signal coefficients are chosen as features. Further, normalization is used for the purpose of enhancing these features. Data normalization is performed to avoid large dynamic ranges in one or more dimensions. We have evaluated the effects of having no normalization; complete normalization of all the elements in the feature vector with respect to the maximum of the individual vector elements; and partial normalization where f_0, f_l and f_u are normalized with respect to 10 kHz and the wavelet features are normalized with respect to the maximum of the individual wavelet features. Table 54.1 summarizes the test conditions evaluated in this work.

54.5 LS-SVM Based Fault Classifier

54.5.1 *Least Squares Support Vector Machine Classifiers*

The support vector machine (SVM) theory developed by Vapnik [21] provides better generalization ability when compared to conventional methods. It uses a kernel function to map the input samples to a higher dimension feature space, where the overlapped samples become linearly separable. This involves solving a complex quadratic programming (QP) problem. The least squares support vector machine [17] reduces the complexity and computations involved in the QP problem.

Given a set of data vectors $(x_1, y_1), \ldots, (x_n, y_n)$, such that $x_i \in \Re^N$ and $y_i \in \{-1, +1\}, i = 1, \ldots, n$, the aim is to construct a classifier of the form $w^T \Gamma(x_i) + b$ subjected to the constraints $y_i(w^T \Gamma(x_i) + b) \geq 1$, where $\Gamma(x)$ is a nonlinear function that maps the input space to a higher dimensional space, where $w$is a M-dimensional weight vector perpendicular to the separating hyperplane, and b is the bias term. In order to include some degree of tolerance in misclassification, a slack variable (ξ_i) is introduced. In simple SVM, the classification problem is given by the saddle point of the Lagrangian function, whose computation leads to the solution of the following QP problem, based on which the classifier needs to be designed

$$\max_{\alpha_i} Q_1(\alpha_i; \Gamma(\mathrm{x}_i)) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j \Gamma(\mathrm{x}_i)^T \Gamma\left(\mathrm{x}_j\right)\alpha_i\alpha_j + \sum_{i=1}^{n}\alpha_i, \tag{54.8}$$

subjected to the constraint $\sum_{i=1}^{n} \alpha_i y_i$ and having Lagrange multiplier $0 \leq \alpha_i \leq C; i = 1, \ldots, n$. To reduce the complexity, the modification introduced by Suykens and Vandewalle [17] is in the cost function:

$$\min_{w,\xi,b} J_2(w, \xi) = \frac{1}{2}\|w\|^2 + \frac{1}{2}\gamma\sum_{i=1}^{n}\xi_i^2. \tag{54.9}$$

Table 54.1 Test conditions evaluated

Condition	Feature type	Wavelet feature type	Normalization
C1	Conventional frequency features only	–	No
C2	Conventional frequency features only	–	All
C3	Conventional frequency features only	–	Partial
C4	Wavelet features only	Energy indicator of detail coefficients $\left(E_j's\right)$	No
C5	Wavelet features only	Energy indicator of detail coefficients $\left(E_j's\right)$	All
C6	Wavelet features only	Mean $\left(\mu_j\right)$ and standard deviation $\left(\sigma_j\right)$ of detail coefficients	No
C7	Wavelet features only	Mean $\left(\mu_j\right)$ and standard deviation $\left(\sigma_j\right)$ of detail coefficients	All
C8	Conventional frequency and wavelet features	Energy indicator of detail coefficients $\left(E_j's\right)$	No
C9	Conventional frequency and wavelet features	Energy indicator of detail coefficients $\left(E_j's\right)$	All
C10	Conventional frequency and wavelet features	Energy indicator of detail coefficients $\left(E_j's\right)$	Partial
C11	Conventional frequency and wavelet features	Mean $\left(\mu_j\right)$ and standard deviation $\left(\sigma_j\right)$ of detail coefficients	No
C12	Conventional frequency andwavelet features	Mean $\left(\mu_j\right)$ and standard deviation $\left(\sigma_j\right)$ of detail coefficients	All
C13	Conventional frequency and wavelet features	Mean $\left(\mu_j\right)$ and standard deviation $\left(\sigma_j\right)$ of detail coefficients	Partial

This transforms the problem to a $(n + 1) \times (n + 1)$ linear system:

$$\begin{bmatrix} \mathbf{\Omega} & \mathbf{Y} \\ \mathbf{Y}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ b \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ 0 \end{bmatrix}, \tag{54.10}$$

where, $\Omega_{ij} = y_i y_j \mathrm{k}\left(\mathbf{x}_i, \mathbf{x}_j\right)$, $\mathrm{Y} = [\mathrm{y}_1, \ldots, \mathrm{y}_n]^T$, $\alpha = [\alpha_1, \ldots, \alpha_n]^T$, and $\mathbf{1} = [1, \ldots, 1]^T$.

54.5.2 Multi-Class SVM

Multiple faults classification problems are typically solved by combining multiple binary classification LS-SVMs [18]. One-against-one LS-SVM (OAO LS-SVM) provides the best balance between the sample numbers of the classes under consideration. OAO LS-SVM for C-class classification constitutes $C(C-1)/2$

classifiers where each classifier is trained by the data from two classes according to the above explained LS-SVM algorithm. Then voting is used for the future testing, after all $C(C-1)/2$ classifiers are constructed, based on the decision function:

$$f_{ij}(x) = w_{ij}^T \Gamma(x) + b_{ij}, \tag{54.11}$$

where, $w_{\{\mathrm{ij}\}}$ and b_{ij} are the weight vector and bias term of the i th and j th classes respectively. Then the data sets are classified into their corresponding classes based on the rule

$$\underset{i=1,\ldots,C}{\arg\max} f_i(\mathbf{x});\ where\ f_i(\mathbf{x}) = \sum_{i=1, j\neq i}^{C} sign\left(f_{ij}(\mathbf{x})\right) \tag{54.12}$$

If the above equation is satisfied for one i, then $\mathbf{x}$ is classified into class i, else if the equation is satisfied for plural i's, then $\mathbf{x}$ is unclassifiable.

54.6 Experimentation and Performance Results

The sample circuit built to examine fault diagnosis in analog filter circuits using LS-SVM is a Sallen-Key band pass filter (Fig. 54.2) having component tolerances ranging within 10 %. The motivation behind selecting this filter circuit is because, it has been the most commonly used circuit for evaluating the performance of fault diagnostic approaches and thus becomes simpler to compare the performance results with the earlier results reported in [10–12], [23–25]. The frequency response of the circuit with $C1$, $C2$, $R2$ and $R3$ varying within their tolerances, belong to the no-fault class (NF). When any of the above four components vary beyond their tolerance range with other components varying within their tolerances, then we obtain faulty responses using which the fault classes are constructed (Table 54.2). These faulty responses along with the NF class responses are used to generate features that will eventually be used to train the one-against-one multiclass LS-SVM classifier.

As shown in Table 54.2, we consider component values that are just (>10 %) higher or lower than the nominal value to values as high as 200 % the nominal value. The experimental setup for evaluating the proposed framework is shown in Fig. 54.3. Once the response signal is obtained, the corresponding signal's power spectrum is estimated using the Welch's method as explained in Sect. 54.3. Then the features are extracted from the spectral estimate. These features are supplied as input to the LS-SVM classifier. One-against-one multiclass LS-SVM was employed in this work, where 36 binary LS-SVM's were trained and optimized. The classifier has been trained using 30 data sets obtained from each intentionally induced fault class. During the diagnostics phase, in each of the nine fault condition, four different types of fault values were used to produce 10 faulty responses each, which was analyzed using the trained LS-SVM classifier. The classifier was

Fig. 54.2 25-kHz Sallen-Key bandpass filter used in our study

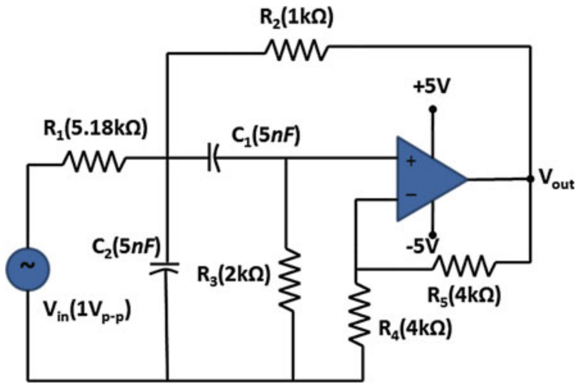


Table 54.2 Fault classes for Sallen-Key band pass filter

Fault class	Nominal value	Faulty value
NF	–	–
*R*2 ⇑(F5)	1 kΩ	1.241 kΩ, 1.613 kΩ, 2.217 kΩ, 3.477 kΩ
*R*2 ⇓(F6)	1 kΩ	43.8 Ω, 372.3 Ω, 669 Ω, 836 Ω
*R*3 ⇑(F1)	2 kΩ	2.311 kΩ, 2.529 kΩ, 3.655 kΩ, 4.447 kΩ
*R*3 ⇓(F2)	2 kΩ	200 Ω, 0.883 Ω, 1.5 kΩ, 1.794 kΩ
*C*1 ⇑(F7)	5 *nF*	6 *nF*, 7 *nF*, 8 *nF*, 10 *nF*
*C*1 ⇓(F8)	5 *nF*	1.5 *nF*, 1.875 *nF*, 2.5 *nF*, 3.5 *nF*
*C*2 ⇑(F3)	5 *nF*	6 *nF*, 7 *nF*, 8 *nF*, 10 *nF*
*C*2 ⇓(F4)	5 *nF*	1.5 *nF*, 1.67 *nF*, 2.5 *nF*, 3 *nF*

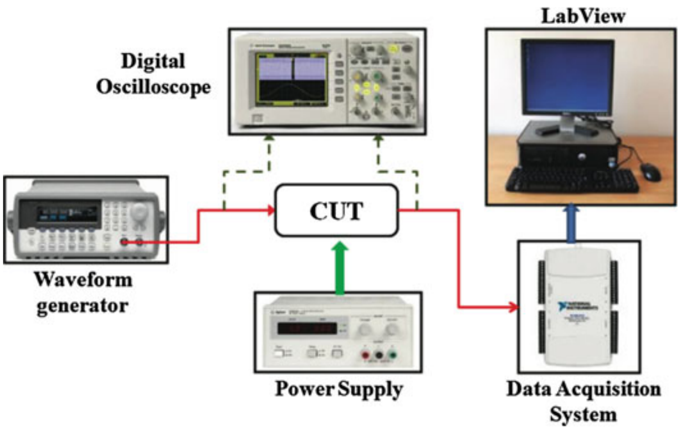


Fig. 54.3 Experimental setup for demonstrating proposed fault diagnostic approach

tested for all the test conditions listed in Table 54.1. Table 54.3 shows, the test accuracy and test time for the designed LS-SVM classifier. In this work, *test accuracy* is defined as the ratio of correctly classified samples to all tested samples; *test time* is defined as the time consumed for testing all of the samples. Among the evaluated test conditions (C1 through C13), test condition C13 outperforms all other test conditions in terms of test accuracy ($\sim$99 %) and test time (around 6.2 s for testing 360 fault samples). From test conditions C10 and C13 (Table 54.3) we can presume that combining conventional frequency features with wavelet features and performing partial normalization always lead to the best classification accuracy. When only conventional frequency features (C1) or wavelet features (C4, C6) are taken into account, wavelet features provided better test accuracy that too in the least testing time. However, if only wavelet features were considered and performing normalization on the feature vectors resulted in poor testing accuracy. Furthermore, by comparing test conditions C4 and C5 with C6 and C7 shows that wavelet features extracted from the statistical properties of the detailed coefficients provides better classification information than the energy contained in the detailed coefficients.

Now, in order to show the significance of the proposed diagnostic framework, we compare our method with the existing diagnostic approaches. Most of the diagnostic approaches presented in the past such as the ones in [3], [8–12], [19, 20], [23–26], considered a CUT to be faulty only when the value of a critical component was higher or lower than the nominal value by 50 %. They do not explicitly demonstrate the performance of their diagnostic approaches when there are small variations in the value of critical component outside the tolerance range. Here, we have demonstrated classification accuracy around 99 % even when the components value varies by 20 % higher/lower than the nominal values. Also, as in the aforementioned methods, if we consider only 50 % variation as faulty condition then our method demonstrates a test accuracy of 100 %.

Table 54.3 Classifier performance for the Sallen-Key band pass filter

Condition	Test accuracy	Test time (s)
C1	0.547	29.281
C2	0.811	14.835
C3	0.969	8.845
C4	0.878	15.397
C5	0.569	38.095
C6	0.936	11.450
C7	0.775	17.300
C8	0.547	28.797
C9	0.833	13.525
C10	0.964	7.706
C11	0.480	29.843
C12	0.908	11.949
C13	**0.989**	**6.193**

In the recently published works of Yuan et al. [24], and Xiao and He [25], 100 % diagnosability has been claimed for the same Sallen-Key band pass filter. However, both these works are based on data collected using simulation study, which are sometimes not consistent with real data extracted from the field. This is because, even though tolerance level can be specified during simulations, simulations do not cover all the values within the standard tolerance range of every single component that naturally arises in an actual circuit. This tolerance problem can obscure the separability of fault classes and thus can affect the performance of the classifier during real life application. As an example, Aminian and Aminian [11] demonstrated a test accuracy of 100 % for a Sallen-Key band pass filter using data collected from a simulation study however, the performance dropped to less than 95 % when experimentally verified [12].

54.7 Conclusions

In this study, we have proposed and experimentally validated a systematic approach to perform soft-fault diagnosis on actual analog filter circuits that are affected by component tolerances. The proposed approach is completely automated and is capable of detecting and isolating faults in real-time and thus, can be used to evaluate the reliability of electronic systems during field operation. The proposed approach overcomes the issues confronted in the existing diagnostic methods. In contrast to diagnostic methods based on circuit nodal equations, the proposed approach is not limited by the inaccessibility to the internal nodes of analog circuits as in modern integrated circuits, as the approach monitors only the response of the circuit at the output node.

The generalization capability of LS-SVM and the use of the structural risk minimization concept ensure that component tolerances do not affect the separability of the extracted features, which is a major drawback in neural network–based diagnostic techniques. Furthermore, the use of LS-SVM transformed the classification problem into a system of linear equations rather than a quadratic programming problem; this helped in reducing the computation time involved in classification. For an analog filter circuit, extracting features from the frequency domain proved to be efficient, as it helped in avoiding additional computations involved with data preprocessing for reducing the number of features. Features extracted using wavelet transformation and performing partial normalization on these features tends to have a significant effect on not only the test accuracy (99 %) but also the testing time (reduced by 50 % e.g., C13-6.193 s/C2-14.835 s, C10-7.706 s/C9-13.525 s, and C3-8.845/C2-14.835 s). Thus, the use of LS-SVM leads to reliable classification in reduced testing time when compared to other previously reported diagnostic approaches. Our proposed approach can be extended to other analog circuits whose performances are defined in the frequency domain. Future work includes the investigation of different mother wavelets, the possibility of detecting and isolating multiple faults which might generate refusal areas during classification using least squares support vector machine, and detecting hard faults.

Acknowledgments The authors would like to thank the more than 100 companies and organizations that support research activities at the Prognostics and Health Management Group within the Center for Advanced Life Cycle Engineering at the University of Maryland annually.

References

1. Birolini A (1997) Quality and reliability of technical systems. Springer, New York
2. Li F, Woo PY (2002) Fault detection for linear analog IC—the method of short-circuits admittance parameters. IEEE Trans Circuits Syst 49(1):105–108
3. Alippi C, Catelani M, Fort A, Mugnaini M (2002) SBT soft fault diagnosis in analog electronic circuits: a sensitivity-based approach by randomized algorithms. IEEE Trans Instrum Meas 51(5):1116–1125
4. Williams A, Taylor F (2006) Electronic filter design handbook. McGraw-Hill, New York
5. Yang C, Tian S, Long B, Chen F (2011) Methods of handling the tolerance and test-points selection problem for analog-circuit fault diagnosis. IEEE Trans Instrum Meas 60(1):176–185
6. Mei H, Hong W, Geng H, Shiyuan Y (2007) Soft fault diagnosis for analog circuits based on slope fault feature and BP neural network. Tsinghua Sci Technol 12(S1):26–31
7. Halgas S (2008) Multiple soft fault diagnosis of nonlinear circuits using the fault dictionary approach. Bull Pol Acad Sci 56(1):53–57
8. Zhou L, Shi Y, Zhao G, Zhang W, Tang H, Su L (2010) Soft-fault diagnosis of analog circuit with tolerance using mathematical programming. J Commun Comp 7(5):50–59
9. Cui J, Wang Y (2011) A novel approach of analog circuit fault diagnosis using support vector machines classifier. Measurement 44:281–289
10. Spina R, Upadhyaya S (1997) Linear circuit fault diagnosis using neuromorphic analyzers. IEEE Trans Circuits Syst II 44:188–196
11. Aminian M, Aminian F (2000) Neural-network based analog-circuit fault diagnosis using wavelet transform as preprocessor. IEEE Trans Circuits Syst II 47(2):151–156
12. Aminian F, Aminian M (2002) Analog fault diagnosis of actual circuits using neural networks. IEEE Trans Instrum Meas 51(3):544–550
13. Mohamed EA, Abdelaziz AY, Mostafa AS (2005) A neural network-based scheme for fault diagnosis of power transformers. Electr Power Syst Res 75(1):29–39
14. Huang J, Hu X, Yang F (2011) Support vector machine with genetic algorithm for machinery fault diagnosis of high voltage circuit breaker. Measurement 44:1018–1027
15. Mao X, Wang L, Li C (2008) SVM classifier for analog fault diagnosis using fractal features. In: Proceedings of the 2nd IEEE international symposium on intelligent information technology and application, pp 553–557
16. Scholkopf B, Smola A (2002) Learning with kernels-support vector machines, regularization, optimization and beyond. MIT Press, Cambridge
17. Suykens J, Vandewalle J (1999) Least squares support vector machine classifiers. Neural Process Lett 9(3):293–300
18. Hsu CW, Lin CJ (2002) A comparison of methods for multi-class support vector machines. IEEE Trans Neural Networks 13(2):415–425
19. Long B, Tian SL, Wang HJ (2008) Least squares support vector machine based analog circuit fault diagnosis using wavelet transform as preprocessor. In: Proceedings of the international conference on communications, circuits and systems, pp 1026–1029

20. Lei Z, Ligang H, Wang Z, Wuchen W (2010) Applying wavelet support vector machine to analog circuit fault diagnosis. In: Proceedings of the 2nd workshop on education, technology and computer science, pp 75–78
21. Vapnik K (1995) The nature of statistical learning theory. Springer, New York
22. Welch P (1967) The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Trans Audio Electroacoust 15(2):70–73
23. Aminian M, Aminian F (2007) A modular fault-diagnostic system for analog electronic circuits using neural networks with wavelet transform as preprocessor. IEEE Trans Instrum Meas 56(5):1546–1554
24. Yuan L, He Y, Huang J, Sun Y (2010) A new neural-network-based fault diagnosis approach for analog circuits by using kurtosis and entropy as a preprocessor. IEEE Trans Instrum Meas 59(3):586–595
25. Xiao Y, He Y (2011) A novel approach for analog fault diagnosis based on neural networks and improved kernel PCA. Neurocomputing 74:1102–1115
26. Toczek W, Zielonko R, Adamczyk A (1998) A method for fault diagnosis of nonlinear electronic circuits. Measurement 24:79–86

Chapter 55
Gearbox Fault Diagnosis Using Two-Dimensional Wavelet Transform

M. R. Hoseini and M. J. Zuo

Abstract In this paper, a novel technique for denoising gearbox vibration has been proposed. We first convert the vibration signal into a two-dimensional matrix such that each row of the resulting matrix contains exactly one revolution of the gear. This matrix is subsequently denoised using two-dimensional wavelet thresholding method. We apply our proposed method to an experimental data set to investigate the improvement in denoising performance. The experimental data is generated using a test rig on which different damage levels are simulated. The experimental results show that the impulses in the vibration signal can be detected easily from the denoised signal even for slight localized tooth damage. The proposed method is compared to time synchronous averaging and the combination of the time synchronous averaging and the one-dimensional wavelet denoising. The kurtosis value of the denoised signal is used for comparing the denoising performance of these three methods. The comparison study shows that the proposed method outperforms both competing methods, especially in early stages of the fault.

55.1 Introduction

Time synchronous averaging (TSA) is a well adopted method to remove noise from the gearbox vibration [1]. In addition to removing noise, TSA separates the vibration signature of an individual gear from the total vibration of the gearbox by removing the non-synchronous vibration. This enables us to detect local gear tooth damages by direct inspection of time domain average. Nevertheless, the average signal has a poor frequency resolution. In addition, TSA relies on the assumption

M. R. Hoseini · M. J. Zuo (✉)
Reliability Research Lab, University of Alberta, Edmonton, Canada
e-mail: ming.zuo@ualberta.ca

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_55,

that the gear vibration is cyclo-stationary however the gear vibration may not be perfectly cyclo-stationary. Moreover, the denoising capacity of TSA is limited by the number of cycles and usually a long run of the signal is required [2]. Therefore, usually TSA is combined with other methods such as wavelet transform to achieve satisfactory denoising performance [3].

As an alternative to TSA method, we propose using two dimensional (2D) wavelet transform (WT) for denoising gearbox vibration. The periodical vibration signal is converted to a two-dimensional matrix. Each row of this matrix contains exactly one revolution of the gear. The main advantage of the proposed denoising method is that it employs the signal correlation within a cycle and across cycles simultaneously, whereas TSA uses only correlation across cycles.

Effectiveness of the proposed denoising technique is demonstrated through a pilot plant case study. On this pilot plant, we have simulated different crack levels on the gear tooth using electro-discharge machining. We apply our denoising method to the vibration signal generated using this test rig. Finally, the proposed method is compared with TSA and a combination of TSA and 1D WT.

The remainder of this paper is organized as follows. Section 55.2 describes the experimental setup. Section 55.3 briefly introduces the wavelet transform. Section 55.4 explains the proposed method. Section 55.5 presents the comparison between the proposed method and TSA and the combination of TSA and 1D wavelet transform. Finally, conclusions are given in Sect. 55.6.

55.2 Experimental Study

Gearbox failures are mostly (60 %) caused by gear faults [4]. The gear faults, in turn, are initiated primarily by the local tooth faults such as pitting, cracking and spalling [4]. In this study, detection of tooth cracks on its early stages has been investigated.

Figure 55.1a shows the gearbox test rig consisting of a motor with adjustable speed, a transmission belt, a two-stage gearbox and a brake to apply the desired load. The number of teeth of each gear and the transmission ratios are given in Fig. 55.1a. All gears are spur gears. Different damage levels have been created on gear #3 as follows: no damage, 25, 50, 75, 100 % crack and missing tooth. The no damage gear has been used to collect baseline vibration. The cracks are simulated by cutting the root of one of the gear teeth at 45 ° by electro-discharge machining. The crack depths are presented as a percentage of the chordal length. For example, the 50 % crack has a depth equal to 50 % of the chordal length. We have increased the crack length proportionally as we have increased the crack depth. The crack lengths are measured as a percentage of the gear width. For example, the 50 % crack has a length equal to 50 % of the crack width. Whenever the crack depth reaches the chordal length (100 % crack), the tooth often breaks rapidly [5]. Therefore, we have removed a tooth completely to simulate the final stage of the crack progress. The 50 % crack and missing tooth gears are shown in Fig. 55.1.

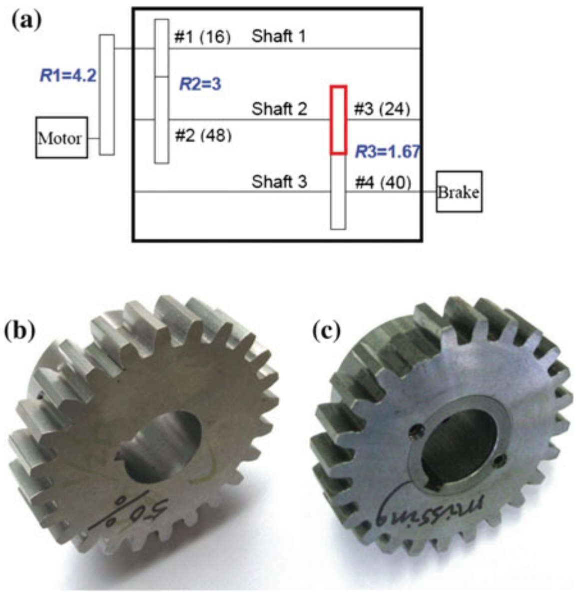


Fig. 55.1 Gearbox test rig configuration (**a**), a gear with 50 % crack on one tooth (**b**), and a gear with a missing tooth (**c**)

The vibration is measured using an accelerometer at sampling frequency of 2,560 Hz at motor speed of 40 Hz.

55.3 Wavelet Transform

Wavelet transform is one of the most powerful tools for feature extraction, signal denoising, and signal compression [6]. WT is localized in both spatial and frequency domains [7]. It also offers more flexibility in selecting the basis than do other transforms such as the Fourier transform. The discrete wavelet transform (DWT) is often performed using the pyramidal algorithm proposed by Mallat [8]. The pyramidal algorithm decomposes a signal, f_m, at level m by convolving it with a low-pass filter, h (scaling filter), to form an approximation signal, f_{m+1}, at level $m+1$, and a high-pass filter, g (wavelet filter), to form a detail signal, f'_{m+1}, at level $m+1$.

$$\begin{aligned} f_{m+1}[n] &= \sum_k h[2n-k]f_m[k] \\ f'_{m+1}[n] &= \sum_k g[2n-k]f_m[k] \end{aligned}. \tag{55.1}$$

The resulting signals are subsequently sampled at every other point in order to avoid redundancy. The 2D wavelet transform can be formulated to be separable so it can be computed by extending the 1D pyramidal algorithm transform. In other

Table 55.1 Frequency content of each wavelet coefficient

		Vertical	
		Low frequency	High frequency
Horizontal	Low frequency	Approximation coefficient	Horizontal detail coefficient
	High frequency	Vertical detail coefficient	Diagonal detail coefficient

words, 1D wavelet decomposition is performed first in one direction, then in the other direction. Thus, the 2D WT of a 2D function, $f_m[n_1,n_2]$, is calculated as:

$$\begin{aligned} f_{m+1}[n] &= \sum_{k_1}\sum_{k_2} h^H[2n_1 - k_1]h^V[2n_2 - k_2]f_m[k_1,k_2] \\ f^H_{m+1}[n] &= \sum_{k_1}\sum_{k_2} h^H[2n_1 - k_1]g^V[2n_2 - k_2]f_m[k_1,k_2] \\ f^V_{m+1}[n] &= \sum_{k_1}\sum_{k_2} g^H[2n_1 - k_1]h^V[2n_2 - k_2]f_m[k_1,k_2] \\ f^D_{m+1}[n] &= \sum_{k_1}\sum_{k_2} g^H[2n_1 - k_1]g^V[2n_2 - k_2]f_m[k_1,k_2] \end{aligned}, \tag{55.2}$$

where superscripts H, V, and D stand for horizontal, vertical, and diagonal coefficients, respectively. Therefore, a 2D signal is decomposed into an approximation signal and three detail signals containing horizontal, vertical, and diagonal details of the signal. Table 55.1 gives the frequency content of each wavelet coefficient. For example, the horizontal detail coefficient contains the horizontal low frequency and vertical high frequency content of the signal.

55.4 Gearbox Vibration Denoising Using 2D WT

To overcome the shortcomings of TSA and to take the advantage of wavelet denoising, we propose a denoising method based on the 2D WT. The 1D vibration signal is converted into a 2D matrix such that each row of the matrix contains exactly one revolution of the gear under examination. Figure 55.2a shows an example of vibration signal collected at missing tooth damage condition. Figure 55.2b illustrates the same signal converted into a 2D image. The damage on the tooth generates an impulse at every cycle and hence we expect to observe a vertical line with high amplitude in the 2D image. However, this fault signature is not manifested clearly due to the presence of random noise and non-synchronous components in the signal. We employ 2D WT to remove both random noise and non-synchronous components.

Figure 55.3 depicts one-level 2D wavelet decomposition of the vibration of the gearbox with missing tooth damage. To perform wavelet decomposition, we have

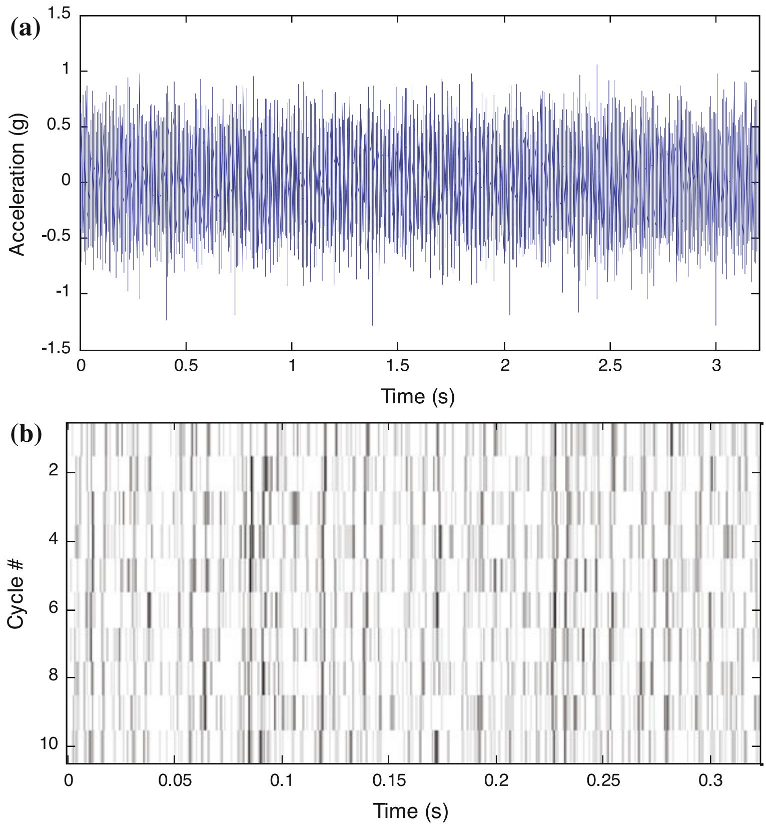


Fig. 55.2 Gearbox vibration for missing tooth damage on gear #3 (**a**), and its 2D representation (**b**)

used Haar wavelet in the vertical direction, because the number of cycles is limited and hence a compact support is desired. The Haar wavelet has the shortest support among all wavelet families. The 'sym 8' wavelet filter has been used as horizontal wavelet filter as suggested by Yuan and Pan [9].

The horizontal and diagonal coefficients mainly contain non-synchronous components, because these components change abruptly from cycle to cycle. Therefore, non-synchronous components can be removed by removing horizontal and diagonal coefficients. The random noise is removed by thresholding the vertical coefficient. To achieve a better denoising performance, the approximation coefficient can be further decomposed. Figure 55.4 shows five cycles of vibration signals denoised using the 4-level wavelet decomposition for (a) a gear with a missing tooth and (b) a gear with a 25 % crack on a tooth. A series of periodic impulses can be seen in Fig. 55.4 which indicates the presence of a gear tooth fault.

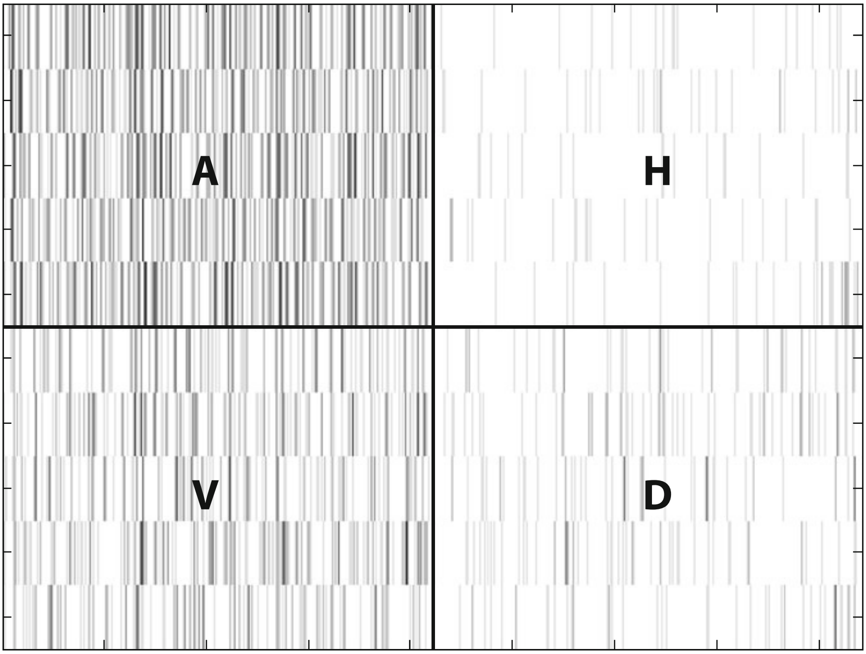


Fig. 55.3 2D Wavelet decomposition of vibration signal of a gearbox with missing tooth damage; *A* Approximation, *H* Horizontal detail, *V* Vertical detail, *D* Diagonal detail. Each coefficient is obtained by filtering and down-sampling of the image illustrated in Fig. 55.2b. Therefore, for each coefficient the x-axis denotes time from 0 to 3.2 s, the same as the Fig. 55.2b and the y-axis denotes the cycle #. However, because of the down-sampling every two cycles are represented by one row

55.5 Comparison with TSA and Combination of TSA and Wavelet Transform

In this section, we compare the proposed method with TSA and combination of TSA and 1D wavelet transform for different damage levels. For comparison of these methods, one cycle of the vibration signals which have been shown in Fig. 55.4 are depicted in Fig. 55.5a and d. Figure 55.5b shows the time synchronous averaged of the vibration of the gearbox with missing tooth damage. Though the impulse, which is caused by missing tooth, may be detected in Fig. 55.5b, it is contaminated by high level noise. This signal is then denoised using 1D WT (Fig. 55.5c). Comparing Fig. 55.5a and c tells us that for the missing tooth fault case, the denoising performance of the combination of TSA and 1D WT is comparable to that of the proposed method.

Figure 55.5d–f show the 25 % crack vibration signal denoised by all three methods. It is hard to detect any impulse in the TSA plot (Fig. 55.5e) as the resulted signal is too noisy. It is observed that the impulse can be detected from the

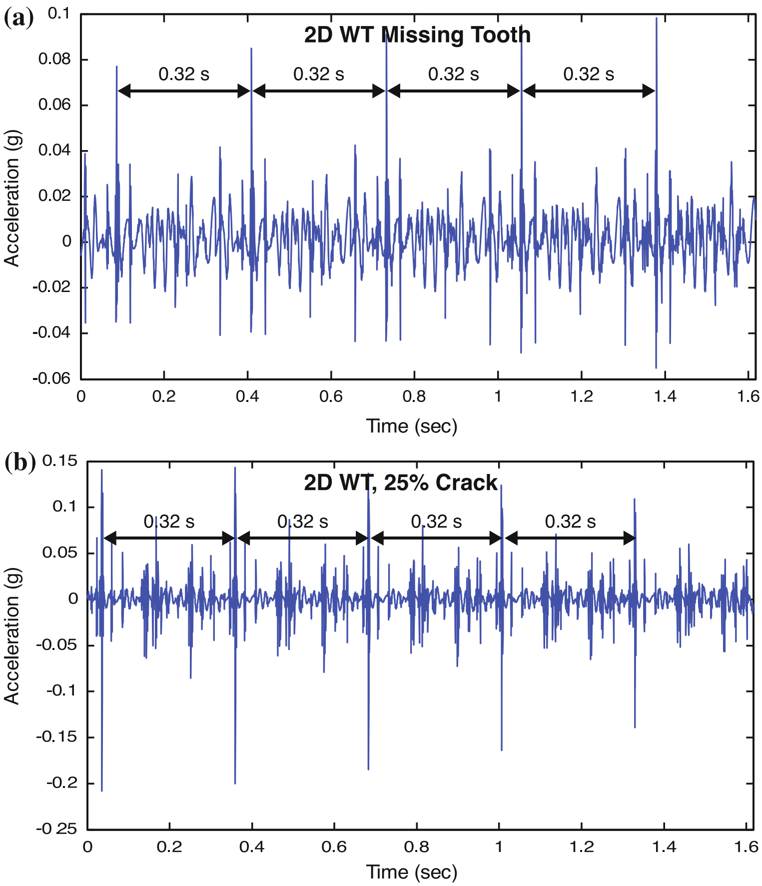


Fig. 55.4 Five cycles of the vibration signals denoised using the proposed method; gear with a missing tooth (**a**), and gear with a 25 % cracked tooth (**b**)

signal denoised by the two other methods. However, the impulse appears more clearly in the signal denoised using the proposed method as the amplitude of the impulse is much larger than the background signal.

In order to quantitatively compare these methods, the kurtosis value of the resulted signal has been computed for all damage levels. Figure 55.6 shows the kurtosis value of the raw vibration signal and the signal denoised by each method. As it can be seen, the kurtosis value for the raw vibration signal is about 3. The kurtosis value of 3 is associated to the normal operation of a system. Similarly, the kurtosis value for the TSA method is also close to 3 and there is only a slight increase from the baseline to the missing tooth damage. This is in agreement with our observation that even for the missing tooth case the impulse in the signal cannot be easily detected in the signal denoised by TSA alone. The combination of TSA and 1D WT performs better than TSA, but still the increase in the kurtosis

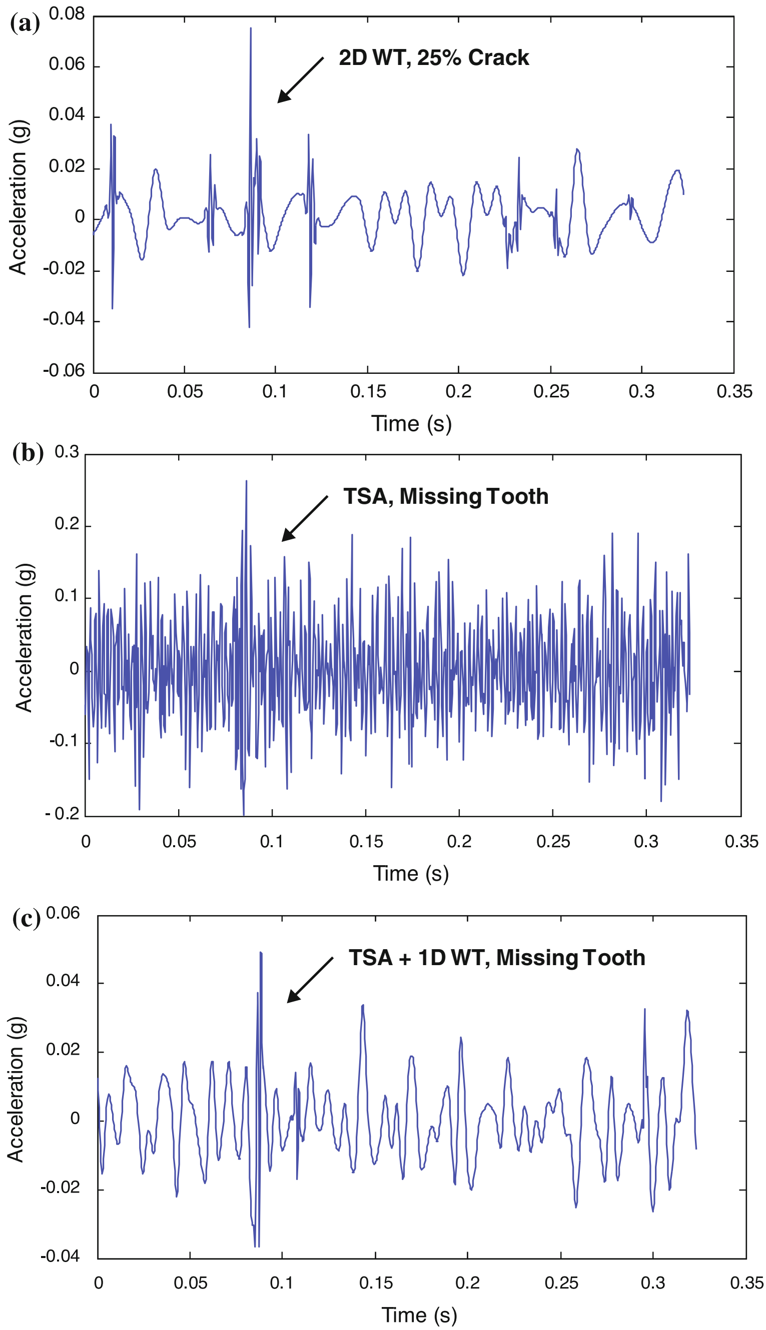


Fig. 55.5 Missing tooth: one cycle of the vibration signal denoised using 2D WT (**a**), TSA (**b**), and TSA and 1D WT (**c**); 25 % crack: one cycle of the vibration signal denoised by 2D WT (**d**), TSA (**e**), and TSA + 1D WT (**f**)

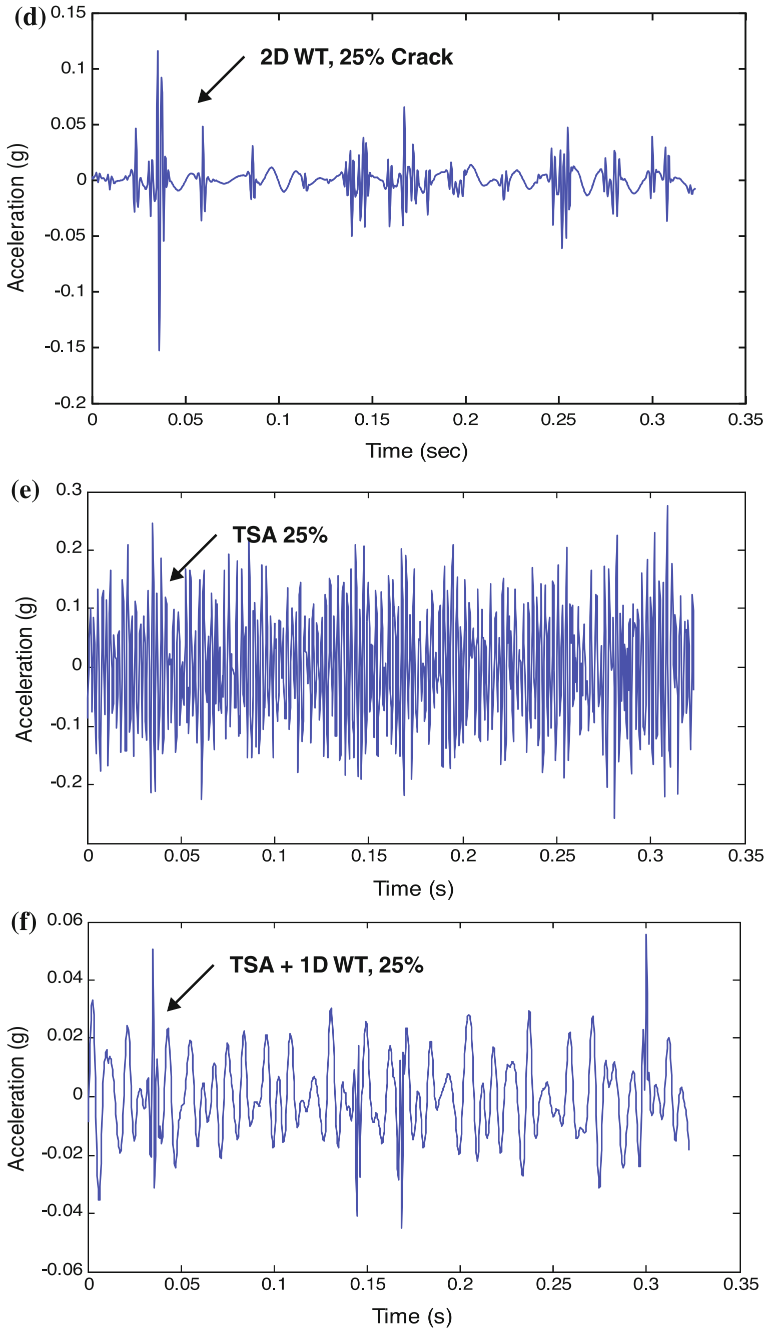


Fig. 55.5 continued

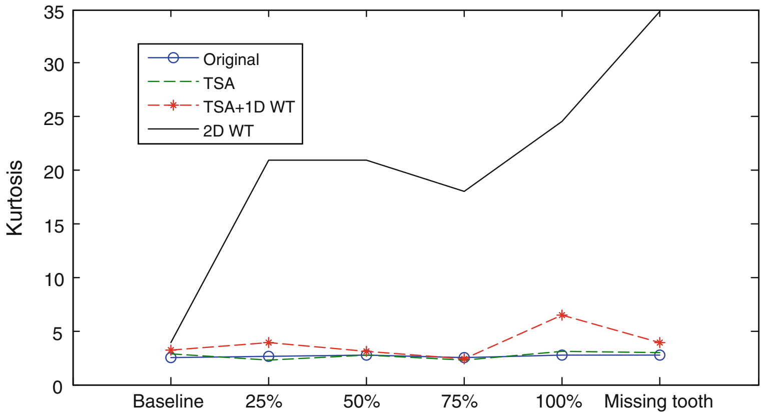


Fig. 55.6 The kurtosis value of the original signal and the denoised signals

value is small. For example, the kurtosis value for 50 % crack is smaller than that of the baseline. Clearly, the proposed method gives the best results and the kurtosis values of damaged cases are much higher than the kurtosis value of the baseline.

55.6 Conclusions

TSA is commonly used for the gearbox vibration denoising. However, the resulted signal has a poor frequency resolution. Its denoising performance is also limited by the number of cycles. To overcome the shortages of TSA, we have proposed a denoising method based on the 2D wavelet transform. The proposed method is able to remove both non-coherent noise and non-synchronous components. An experimental study has been conducted to investigate on the applicability of the proposed method. Different crack sizes have been created on a gear and the vibration at each damage level has been measured. The comparison between our proposed method and TSA and the combination of TSA and 1D WT shows that our method outperforms the competing methods. In this paper, we have mainly focused on the concept of the proposed method. In future work, we will investigate different aspects of the proposed method such as the theoretical noise attenuation capability of the proposed method and the optimal selection of wavelet functions.

References

1. McFadden PD, Smith JD (1985) A signal processing technique for detecting local defects in a gear from the signal average of the vibration. In: Proceedings of the institution of mechanical engineers, vol 199(C4). pp 287–292

2. Hochmann D, Sadok M (2004) Theory of synchronous averaging. In: Proceedings of IEEE aerospace conference; 6–13 March; Big Sky, MO. vol 6. pp 3636–3653
3. Halim EB, Choudhury MAAS, Shah SL, Zuo MJ (2008) Time domain averaging across all scales: a novel method for detection of gearbox faults. Mech Syst Signal Process 22(2):261–278
4. Birkel J, Gelman L, Leigh-Firbank H, Simms D, Waterland B, Whitehurst G, Zimroz R (2005) Adaptive vibration condition monitoring technology for local tooth damage in gearboxes. Insight Nondestr Test Condition Monit 47(8):461–464
5. Tian X (2004) Dynamic simulation for system response of gearbox including localized gear faults. [MS Thesis]. University of Alberta, Edmonton, Canada
6. Mallat SG (1999) A wavelet tour of signal processing, 2nd edn. Academic Press, New York
7. Wang J, Huang HK (2009) Handbook of medical image processing and analysis: three-dimensional image compression with wavelet transforms, 2nd edn. Academic Press, New York
8. Mallat SG (1989) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell 11(7):674–693
9. Yuan J, Pan H (2011) Gearbox vibration signal online de-noise method based on regularized particle filter. J Comput Inf Syst 7(3):786–793

Chapter 56
Flexible Gas Infrastructures

P. Herder and K. Pulles

Abstract The future natural gas network will deliver multiple services and the gas in the network will be a mix of various gasses like hydrogen and biogas. In our study we use asset management techniques and processes to create a flexible, future-proof infrastructure. Ultimately, we will develop a framework in which technical and institutional flexibility are considered. Our first modeling efforts led to a representation of such a changing gas grid, including its physical and operational performance. We used this model to execute some initial Monte Carlo simulations. This first exploratory model allowed us to demonstrate the approach and to explore the value of flexibility in gas infrastructure asset management.

56.1 Introduction

Energy provision in the Netherlands is to a large extent based on natural gas. To this end, a finely meshed natural gas grid is present in the Netherlands, practically reaching almost every household or company. The main source of natural gas is the large natural gas source in the north of the country. The future gas network will be different from the current operating network as sources are depleting and a move toward green energy is being pushed. The future gas network will therefore need to deliver multiple services and the gas in the network will be a mix of various gasses like hydrogen and biogas. The current unidirectional, uni-gas network is undergoing a transition toward a network for sustainable energy sources. As a consequence, the future gas distribution and transport infrastructure will face the challenge of coping with the required increased flexibility of gas production

P. Herder (✉) · K. Pulles
Faculty of Technology, Policy and Management, Delft University
of Technology, Delft, The Netherlands
e-mail: p.m.herder@tudelft.nl

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_56,

(for example, distributed biogas production is less reliable than the current central sources) and the increased variability of gas quality (for example, in terms of calorific value). The current infrastructure needs to be able to accommodate this variety of gas qualities and quantities without compromising the current reliability and safety of our energy provision.

56.2 Asset Management

The imminent transition of the gas sector will not only influence the design of the infrastructures and its components, but also the operations and maintenance of the assets involved. Asset management comprises the design, planning, operation, maintenance and disposal of the physical assets in such a way that the asset base supports optimal performance of the organization. Most of the assets that will be used for the future systems are currently already in use (their replacement will take decades) so the transition to multigas systems will require an adaptation of the asset management practices for the current gas infrastructure and a reviewing of the existing asset management strategies. It will no longer be sufficient to maintain the current functionality of the infrastructure, but it will be necessary to actively integrate new demands with respect to flexibility and quality variability of the multi-gas options. New investments in innovative physical equipment and adaptations to existing equipment will be necessary that will allow for flexibility in the infrastructure's future, for example smart meters, more advanced gas quality measurements, dynamic net control devices, local gas storage options or flexible biogas injection stations.

56.3 Risks and Real Options

Current asset management practice by many (Dutch) grid operators is rooted in a risk-based attitude on the existing mono-gas situation. Risks are generally seen as 'downside' risks, i.e. a risk event would negatively influence system performance, like system failures, leaking pipes, failing control mechanisms, but also non-technical risks such as regulatory risk, public image, etc. [1]. Risk management strategies and tools are often embedded in the asset management processes [2, 3].

However, the future also poses opportunities ('upside risk') i.e. a risk event that could positively influence system performance, like improved tariff structures, increased gas consumption, better than expected technical life time of the assets, etc. Especially when considering future systemic changes to the infrastructure system, such as the mixing in of other gasses, decentralization, and dynamic net management, real options thinking is relevant. It may prevent lock-in into the current infrastructure system, by spotting opportunities and valuing those in a lifespan perspective. Preparing a system consciously for upside risks is known as Real Options Analysis (ROA). It has been shown [4, 5] that real option thinking increases the value of the system by making it more flexible and thus better prepared for an uncertain future.

As grid companies have to justify their current investments to the national regulator, their positions with regard to real options have to be understood by the regulator. Otherwise, extra investments that show no immediate or guaranteed payoff (in terms of financial payoff or keeping the system within legal constraints) would be considered wasteful by a regulator. The advantage of the real options approach is its transparency and financial underpinning of the required "extra" investment that will ultimately lead to overall lower societal cost, thereby safeguarding at least the public value of overall low cost service provision (affordability). This idea is not yet incorporated into the regulatory framework and it has been shown that some barriers still need to be overcome before a practical implementation of the ROA approach is achieved even within the sector [6, 7]. Such barriers range from lack of (lifespan) data, lack of practical methods to execute the seemingly complex ROA calculations, to organizational and regulatory constraints, such as a strict division in many companies between the investment or project organization and the operational divisions.

Another important issue that is still underexplored is how to monitor the status of the real options that have been embedded in a system [8]. Methods for monitoring a changing environment have been developed for policy making, but those notions have not found their way yet to the engineering community [9–11]. The need for monitoring the status and value of the infrastructure will expectedly profit from developments regarding high tech gas network monitoring instruments, such as new sensors and inspection techniques. How to integrate the information obtained with these potential methods into the asset management process is, however, still an open issue.

56.4 Research Hypothesis

As the quality, and the future development and evolution of the physical infrastructure is to a large extent shaped by asset management, we hypothesize that it must be used to nudge the system toward achieving a sustainable, multigas infrastructure. Therefore, we have started research into how real option thinking can increase the future performance of the system, in terms of its flexibility and sustainability. This research will consist of a further extension of the ROA theory and its implementation with regard to the issues mentioned. It will be based on approximately three real-life case studies undertaken in cooperation with several Dutch Distribution System Operators.

56.5 Monte Carlo Model Setup

A first result of the research is the development of a basic ROA-in-asset-management-tool. The tool is an implementation of a Monte Carlo simulation engine which integrates a stochastic demand model with a gas grid design and dimensioning algorithm, using rigorous physical/thermodynamic modeling of the gas distribution grid.

A gas distribution system in the simplest version of the model consists of:

- demands (positive demands of conventional customers and negative demands from biogas suppliers)
- pipeline routes (with or without the actual pipes)
- station locations (from which the distribution grid can be fed, with or without the actual installations).

The model captures the economic/technical issues of the gas distribution network and the business model of the network owner. The assets and demands are categorized into a limited number of definitions. There are separate definitions for the demands, the pipelines and the stations.

In addition to its definition each item has a few additional data:

- a name
- a location (demand and station)
- an optional, fixed commissioning date
- an optional, fixed decommissioning date
- a route (pipeline)
- a length (pipeline)
- an outlet pressure (station).

The estimation of the demand is the stochastic part of the model. Each potential customer (demand) is described by the following parameters:

- initial date (starting date for commissioning probability)
- stochastic time delay for commissioning, described by the rate function of the Weibull distribution (as this distribution fits the degrading curves of for example boilers and heaters very well. The biogas installations coming online are less well represented by Weibull distributions, but the model can be easily adapted to incorporate more befitting stochastic distributions:

$$p_c(t) = k_c/t_c(t/t_c)^{kc-1}$$

where:

- p_c: time-dependent probability-density for delay to commissioning
- t_c: characteristic time delay to commissioning
- k_c: parameter, typically ≥ 1, shape factor

- stochastic time delay for decommissioning, described by the rate function of the Weibull distribution: $p_d(t) = k_d/t_d(t/t_d)^{kd-1}$
  where:
  - p_d: time-dependent probability-density for delay to decommissioning
  - t_d: characteristic time delay to decommissioning
  - k_d: parameter, typically ≥ 1, shape factor
- a set of zero or more other customers, each which can block the commissioning
- a set of zero or more other customers, all needed to be commissioned before commissioning
- a set of zero or more other customers, whose decommissioning triggers commissioning
- optional fixed commissioning date (when set overrules the stochastic model)
- optional fixed decommissioning date (when set overrules the stochastic model).

Each demand is associated with a demand definition, which prescribes the requested capacity (as function of ambient temperature) and yearly revenue for the network owner. The income of the network owner is based solely on the capacity of the actual connection. Note that, due to unbundling in the energy sector, the natural gas itself is traded by another company. The revenue is income for the grid owner between years of commissioning and decommissioning of the demand, if it succeeds in connecting the customer with pipelines and stations with sufficient capacity.

Each distribution station can be considered as an option to construct a station at a given location at a certain time and maintain it for a certain period, not exceeding its technical lifetime. A station is defined by a station definition, which consists of a capacity, lifetime and construction cost and yearly maintenance cost.

Each pipeline can be considered as an option to construct a pipeline in a given route (length) and maintain it for a certain period, not exceeding its technical lifetime. A pipeline is further defined by a pipeline definition, which consists of a diameter/material, lifetime, capacity and construction cost and yearly maintenance cost (per unit length).

56.6 Example

A minimum network consists of a single station, some pipes and some customers. See Fig. 56.1. Not visible in the figure are a modeled demand at the position of DistrictA and another district station at the positions of DS_Mainstreet. This additional demand is assumed to be commissioned when the original demand (DistrictA) is decommissioned. The additional station is commissioned when the life time of the original station is exceeded and there is still gas demand (either by DistrictA or by District_Mainstreet).

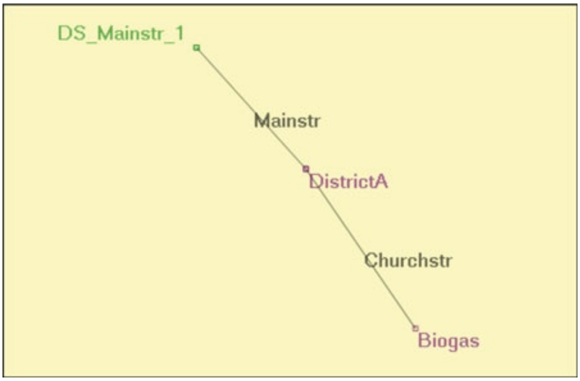


Fig. 56.1 Layout of the minimum network

As an example, two stochastic realizations (or instantiations) of DistrictA (with $t_c = 3$ years, $k_c = 1$, $t_d = 40$ years, $k_d = 1$) are shown in Fig. 56.2.

The revenues are calculated as present values, at the starting time of the simulation, at an interest rate of 5 %. The yearly revenue for this type of demands is set at €11,000 per 100 residential customers. This amounts to a total revenue of € 143,181 and €118,195 in the examples.

The Monte Carlo simulation explores multiple realizations. The demand distribution of 10,000 of these realizations is shown in next figure (scaling by 1,000) (Fig. 56.3).

A histogram of the revenues, based on 10,000 realizations is shown in Fig. 56.4.

Of course, the network owner has to invest, in order to provide capacity for the demand. A very inflexible strategy would be the investment in the construction of the station and the two pipelines at the starting date of the simulation. This would result in a fixed cost in the first year (or for some fixed interval if the costs would be financed over a certain period).

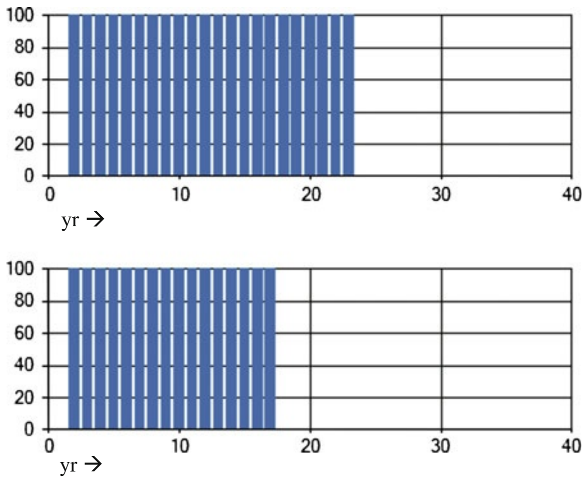


Fig. 56.2 Typical realizations of the demand of DistrictA (vertical axis: demand in m^3h)

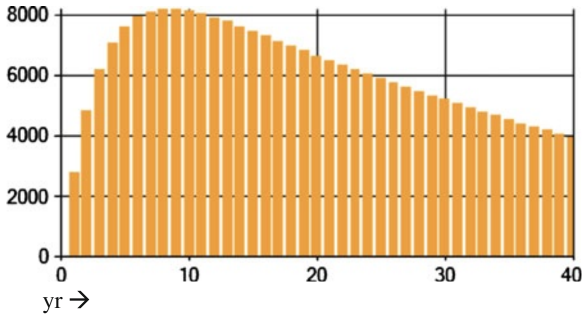


Fig. 56.3 Average demand based on 10,000 realizations (vertical axis: demand in m^3h, scaled by 100)

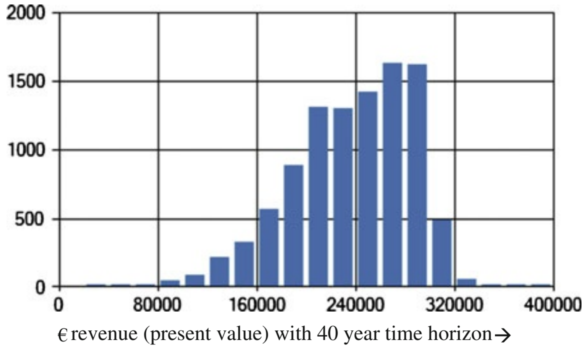


Fig. 56.4 Histogram of revenues based on 10,000 realizations

A more flexible strategy would be the construction of the station and the pipeline(s) just in time (as the demand arises). We assume that the capacity of the station and the pipelines are sufficient for both demands. A histogram of the costs of this strategy is shown in Fig. 56.5.

A histogram of the profits (some negative in this case) of this strategy is shown in the next Fig. 56.6.

The average revenue over the 10,000 simulations is €237,933 and the average cost is €189,299, which results in an average profit of €48,634 over the 40-year simulation period as present value.

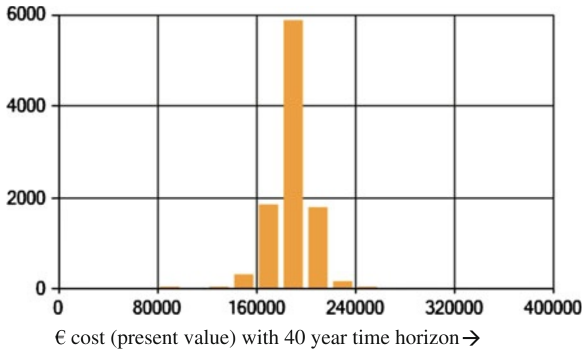


Fig. 56.5 Histogram of total costs in 40 years, based on 10,000 realizations, using a "just in time" strategy

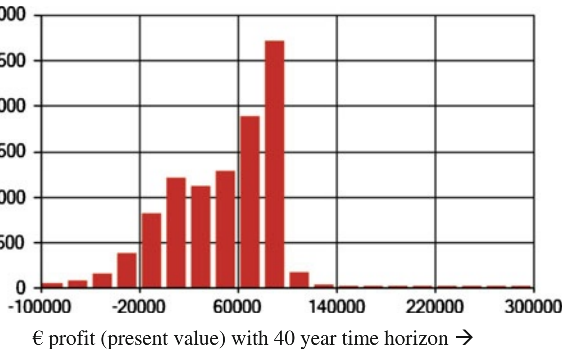


Fig. 56.6 Histogram of total costs in 40 years, based on 10,000 realizations, using a "just in time" strategy

The losses occur mainly in those scenarios where the demand is relatively quickly decommissioned.

56.7 Conclusions and Future Work

The work up to now has been of an exploratory character, demonstrating the feasibility of the implementation. Various refinements of the model will be added in the course of the investigation. For instance, the dimensions of the piping should not only be based on the demand as expected at minimum (winter) temperature, but also on the demand during summer time. This is especially relevant when a continuous supply of biogas is present. Also, the system must still be able to deliver the gas at sufficient pressure to the conventional customers in the absence of the biogas supply.

More types of assets and a more flexible use of the assets could be contemplated and build into the simulation. Under circumstances, active control of the outlet pressure of the regulating stations could be an economical alternative for replacing or adding pipelines by pipelines with other diameters. Gas buffering in summer time could be an alternative for the shutdown of bio gas production.

Finally, more sophisticated strategies should be formulated and tested. The "just in time" strategy should be tested against conditional look ahead strategies i.e., the dimensioning of the grid should not be based on the required capacity of the next year, but based on the expectations for the next decades, taken into account the realization of the demands at a given point in time.

Acknowledgments This research is part of a project in the Dutch research program on gas, EDGaR. Funding by the Dutch DSO's Alliander, Enexis and Stedin is greatly acknowledged. This draft paper is a slightly adapted version of a paper that has been submitted to the International Gas union Research Conference 2011. After review the final submission to WCEAM2011 will incorporate the most recent findings.

References

1. Renn O, Klinke A (2002) A new approach to risk evaluation and management: risk-based precaution-based and discourse-based strategies. Risk Anal 22(6):1071–1094
2. Masera M, Wijnia Ype, Vries Laurens, Kuenzi Caroline (2006) Governing risks in the European critical electricity infrastructure. In critical infrastructures at risk. Top Saf Risk Reliab Qual 9:117–152
3. Wijnia Y, Herder PM (2005) Options for real options—dealing with uncertainty in investment decisions for electricity networks. In: Proceedings of 2005 IEEE international conference on systems, man and cybernetics, October 2005 Hawaii, USA
4. de Neufville R (2003) Real options: dealing with uncertainty in systems planning and design. Integr Assess 4(1):26–34
5. Kulatilaka N (1993) The value of flexibility: the case of a dual-fuel industrial steam boiler. Financ Manage 22(3):271–280
6. Herder PM, Joode J de, Ligtvoet A, Schenk S, Taneja P (2011) Buying real options–valuing uncertainty in infrastructure planning futures. 43(9):961–969
7. Lander DM, Pinches GE (1998) Challenges to the practical implementation of modeling and valuing real options. Q Rev Econ Financ 38(3):537–567
8. Neufville R de, Scholtes S (2011) Flexibility in design. MIT Press. Forthcoming
9. Dewar JA, Builder CH, Hix WM (1993) Assumption-based planning; a planning tool for very uncertain times. Rand Corporation, Santa Monica
10. Dewar JA (2002) Assumption-based planning: a tool for reducing avoidable surprises. Cambridge University Press, Cambridge
11. Walker WE, Rahman SA, Cave J (2001) Adaptive policies, policy analysis, and policy-making. Eur J Oper Res 128(2):282–289

Chapter 57
The Impact of Innovative Contracting on Asset Management of Public Infrastructure Networks

L. Volker, M. Altamirano, P. Herder and T. van der Lei

Abstract The introduction of innovative contracts caused a shift in the roles and responsibilities between asset owner, asset manager, and service providers. This shift requires better integration of decisions on strategic, tactical, and operational level. In this paper we discuss the impacts of social technical trends on the asset management of public infrastructure networks. We identify trends in literature and illustrate some of the trade-offs for road maintenance. We conclude by identifying the effect of innovative contracting and other trends on the implementation of asset management systems and proposing a new direction for research into more dynamic contracting.

57.1 Introduction

Asset management is a fast growing field, not only in the context of manufacturing but also in public works. Since World War II many new structures have been built. Now, some 50–60 years later we see a steady increase in maintenance expenditure which clashes with the need to control governmental expenditure. Compared to

L. Volker (✉) · P. Herder · T. van der Lei
Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, Delft, The Netherlands
e-mail: l.volker@tudelft.nl

P. Herder
e-mail: p.m.herder@tudelft.nl

T. van der Lei
e-mail: t.e.vanderlei@tudelft.nl

M. Altamirano
Deltares, Rotterdamseweg 185, Delft, The Netherlands
e-mail: m.a.altamirano@tudelft.nl

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_57,

Table 57.1 Main characteristics of infrastructure and implications for the asset management system

Characteristics	Implication for system
A very long lifespan	Uncertainty about future directions
No resale value	Physical infrastructure is hardly ever removed
Passive elements	Robust and flexible design needed
Built in agile conditions	Lack of administration of current status
Evolutionary systems	Path dependency and lock-in
Networked systems	Robustness high
Widely distributed	Hard to identify all risks
Owners, manager and operators usually not users	Different interests and responsibilities
Users often anonymous	Very little control over use

mechanical equipment, the speed of deterioration of civil structures like roads, dams, bridges, and dikes, is often much smaller [1]. Yet the rate of economic obsolescence is also much lower, implying that the maintenance cost of the structures over the lifecycle is still very substantial. Table 57.1 gives an overview of the main characteristics of infrastructure and implications for the asset management system (based on [2]). This means that managing infrastructure assets often faces complex uncertainties, often more complex and deep that those found in other forms of asset management.

Especially in the public sector asset management is getting more and more attention. This has to with the reform agendas of the government. The Dutch Next Generation Infrastructure foundation (www.nextgenerationinfrastructure.eu) identifies several trends in her positioning paper on asset management:

- from object related maintenance management to life cycle thinking and asset management systems
- split between organizations of asset owners, asset managers, and service providers
- greater need for risk based management
- increasing interdependency of network and maintenance activities
- increase of the use of innovative outsourcing contracts with service managers and providers
- more attention for public values and user participation

Maximizing value and minimizing risks are important drivers for optimization of the asset portfolio and system. Asset management systems, such as PAS 55, distinguish different levels (e.g. Strategic, Tactical and Operational Level; Policy, Strategy, Objectives, and Plan) and actors (e.g. Asset owner, Asset Manager, and Service provider) to be included in asset management. According to Malano et al. [3] general principles and functions of asset management for irrigation and drainage infrastructure include asset planning and creation strategies, operation and maintenance, performance monitoring, accounting and economics, and audit and renewal analysis. Asset management therefore includes maintenance, renovation as well as reconstruction of a variety of assets. Moon et al. [4] identify policy-driven, performance-based, trade-off identification, quality-based, and performance monitoring as five requisite attributes of a transportation asset management system. Therefore system-level performance measures, models, interoperable databases should be used by asset groups to make evidence-based decisions.

Civil networks like the transportation network are important public infrastructures in many countries as they enable economic activity. This network can be considered a socio-technical system [5, 6]. According to Dekker and Scarf [1] optimization of the civil maintenance area proved to be successful because of the following aspects: (1) the problem is repetitive, the problem structure remains more or less the same, and the large duplication of the parts to be maintained (lane sections) allows for structured data collection; (2) a lot of money is involved, hence there is a need for better decision making tools; (3) the problem owner is open about the problem (no competition) and faces an allocation problem which requires objective methods; and (4) the management and execution of maintenance are separated, which allows management to take a much longer term view.

Since the evaluation of Dekker and Scarf in 1998 we see a considerable change in the market structure, the budgets and stakeholder interest that affects the issue of openness about the problem and the objectiveness of the allocation problem. Currently the integrated and performance based contracts seem to be preferred over the traditional Design-Bid-Build (DBB) contracts based on specific requirements. This development is accompanied by an increase of transaction costs and room for opportunistic behavior from agents/contractors. And therefore an increasing need for reliable, accurate, and complete data. Despite the expectations, often these innovative contracts have not led to higher levels of flexibility in the contractual relationship and increased trust between asset managers (principals) and service providers (agents). In the paper we show the complexity of these changes in order to propose strategic solution approaches. We focus on the context of European road networks, formulate key phases for procuring asset management contracts and conclude with directions for further research that we are planning.

57.2 Procurement and Contracting Trends: New Governance Modes

Four trends can be recognized in contracting and procurement of road infrastructures [7]: combined or integrated contracts, a shift from technical requirements to functional or even performance based specifications, indirect financing of projects, long term contracts and alternative awarding criteria besides price. These innovative practices have caused a shift in governance from a public bureau to a market-like arrangement. In the old situation the public road authority did outsource but did so by tendering work orders. Contracts were signed directly with

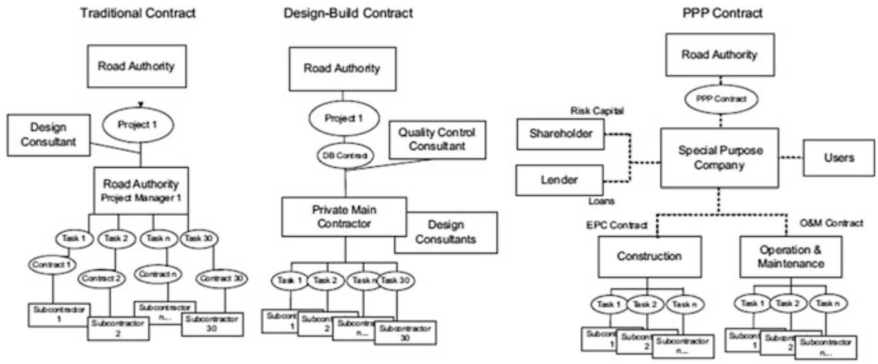


Fig. 57.1 Old versus new organization of contracting entities

each contractor. The road authority had its own staff looking after the management and coordination of all project activities. Private parties only acted as a hired hand in the public sector [8]. In the new situation, the road agency signs a single contract for an entire project or a service agreement for a particular network for a particular time-span. Subcontracting of specialized works or particular work orders still occurs, but now under the responsibility and supervision of the main contractor. The new role of the private contractors is that of a service provider (see Fig. 57.1).

The first column shows a traditional design-bid-build (DBB) project, whereas the second represents the common structure for a design-build (DB) project. In a DB contract, design and construction aspects are contracted for with a single entity known as the main contractor, usually the general contractor; but in some cases it is also the design consultant. Where the design-builder is the main contractor, the design consultants are typically hired directly by the contractor. By making both phases overlap the road agency or infrastructure owner aims at minimizing project risks and delivery times—by allowing the construction contractor to bring his construction expertize into the design process cutting the extra tendering process for the design phase and realizing payments upon completion. Besides functional requirements, instead of technical requirements, are used in combination with lump-sum payments—an up-front negotiated total price for the whole project, instead of unit prices—ensuring the owner against cost variations. Integration of different life cycle phases and use of functional requirements in contracts entail larger design space for contractors, freedom to choose different technical solutions. The third column is illustrative of a public–private partnership (PPP) or concession contract. All life cycle phases of the infrastructure are tendered in a single package.

Before the reform was initiated by all countries, the system was designed under the principle that all operational risks and the entire responsibility toward users were to be fully and directly borne by the public authority. Innovative contracts on periodic maintenance are expected to broaden contractors decision rights and freedom to its maximum level; a contracting setting where the contractor itself decide which road section, when and what kind of work he will perform, with the only condition of guaranteeing a certain level of performance for the whole road network assigned to him.

57.3 Trade-offs in Road Network Maintenance

If innovative contracting practices become the norm, trade-offs about asset management decisions will be transferred to private contractors or other kinds of service providers. This section presents an illustration of the many trade-offs and decisions that are nowadays realized by public infrastructure operators by using a causal loop diagram (CLD) [7]. CLDs are visual representations of cause-effect relationships in a system and a major component and the most commonly used element of the systems thinking approach [9]. The analysis the dynamic behavior of the system based on CLDs is conducted by identifying all feedback loops present in the system. Feedback loops can be reinforcing or balancing. Reinforcing loops reflect positive feedback systems. They can represent exponential growth or declining behavior over time. Unlike reinforcing loops, balancing loops reflect negative feedback systems and seek stability or return to control. The interaction and interdependencies between these loops influence the overall dynamics behavior of the system.

In this causal loop diagram (Fig. 57.2) the two opposite effects of winter tire condition (grip or skid) on safety are shown, represented by number of accidents: First, a direct positive effect—more important during winter—by increasing the total friction between tires and road, which makes easier for drivers to keep control of their cars and therefore reduces the chances of accidents. Second, a negative effect—more important during summer—due to the acceleration they cause in the deterioration rate of the road, through the phenomenon of Wear and Tear. The heavier the tire skids the more the rutting deterioration rate and the actual rut depth. The deeper the ruts on the road, more water can be accumulated (in case of rainy road circumstances) increasing the chances of hydroplaning accidents. It is therefore important to weigh both effects and find the range that is best for reducing both kind of accidents, or that allows the most cost effective maintenance of the road. More details about road condition variables and the trade-off a qualitative system analysis of road network condition for a cold climate country that has been conducted on this illustration, see Altamirano et al. [10].

The factors presented on this CLD correspond to the different aspects of the socio technical system [2]. For instance:

- *Performance requirements* are defined for the different condition related variables (actual road macro texture, rut depth, etc.) and in order to plan needed maintenance activities,
- *Data* is periodically gathered about these road condition characteristics,

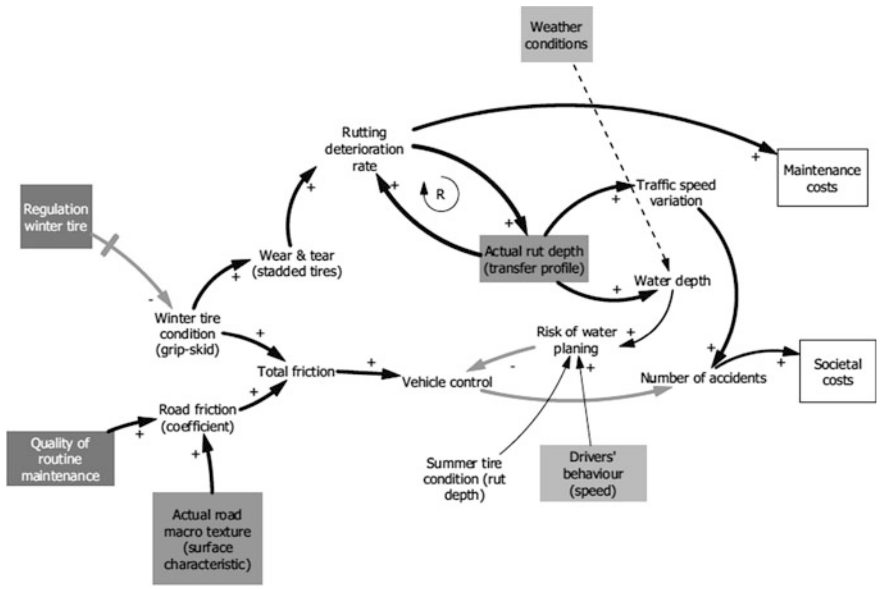


Fig. 57.2 CLD of winter tire condition versus rut depth [7]

- While *public acceptance* set the limits for variables such as number of accidents and other societal costs.
- The assigned *budget* set constraints to the total of maintenance costs, and
- There are national *legal requirements* concerning the use of winter tires in the winter period, meanwhile
- A number of *habits* from the side of road users—like their driving behavior in terms of speed—also affect the performance of the road network.

The boundaries of the asset management system can be broadened to explore the impact of other policies in the friction related accidents. By doing so we discover that total friction can be increased also through a higher service level of winter (routine) maintenance or improving the macro texture of the road section. Therefore it is important to research the maximum levels possible of these different variables and their behavior in terms of costs (e.g. it is achieving the last 5 % much more expensive than the first 95 %?). Once information about these different aspects is known, better informed decision-making could be achieved, a process that will include for example the following possibility: Could it be invested extra in winter maintenance actions what is now invested in the repair of ruts and achieve in this way a much better performance in terms of safety and costs?

Important to notice in this illustrative CLD is the reinforcing loop that takes place in the rutting process (the more the actual rut depth, the bigger the rutting deterioration rate and the higher this, the more the actual rut depth will be) and which could provoke a fast and exponential process of deterioration. Such reinforcing processes do not seem to take place in the area of routine maintenance. This could mean that one could expect an investment in the reduction in the rate of rut deterioration to be more effective or have a higher impact in the overall performance of the system (in terms of safety) than other measures with the same cost. In other words, it could be more effective to try to increase friction through other ways that do not activate or worsen the rutting deterioration rate as the winter tire grip does. It seems also then that improving winter maintenance is an easier point of influence, since this does not show side effects, as winter tire skids do. Here the questions would be, can we really improve the quality or winter maintenance much further? Other aspects to be noticed are the factors dependent on road users *habits*—that could also be a way to solve the whole issue of safety and risk of hydroplaning. Here only two are considered (1) Driver's behavior, mainly speed, and (2) Summer tire condition.

The trade-offs presented here is only meant as an example of the kind of operational dilemmas faced by public network administrators; in this case especially relevant for authorities in charge of road networks exposed to extreme weather conditions. Other dilemmas, faced for instance by road authorities in charge of highly congested road networks are for example to choose for a reduction in the inconveniences caused to road users (which minimizes societal costs) by promoting the realization of maintenance works during nighttimes or weekends or to choose for lower costs of maintenance works (which minimize the authority own costs); or to choose for the most convenient moment for the realization of maintenance (realizing maintenance works precisely when needed) or for a good spread of the workload for the convenience of own personnel (reducing the chance of under or overload) [11].

Besides, there are a number of operational dilemmas valid for network administrators under all climate or traffic conditions, such as to choose for faster and cheaper construction of roads or for reduction of total life cycle costs or total costs of ownership or to choose for extra investing on projects in order to become a society wide example in the area of sustainability and respect for the environment, or focus simply on fulfilling the minimum legal obligations. All the trade-offs need further thinking and society wide discussion if the idea is to transfer them to private agents. The issue of users' satisfaction deserves special attention since there seems to be a trend to make indicators related with it more decisive for the payment received by contractors.

57.4 Ways to Improve Contracting Practice

Looking at organizational level of asset owners and asset managers, the changes in the socio technical system should be visible in the development of public organizations and their market approach. Schraven et al. [12] found in a case study about a Dutch provincial road agency that the alignment between three decision areas (objectives and strategic goals, technical and functional performance-related

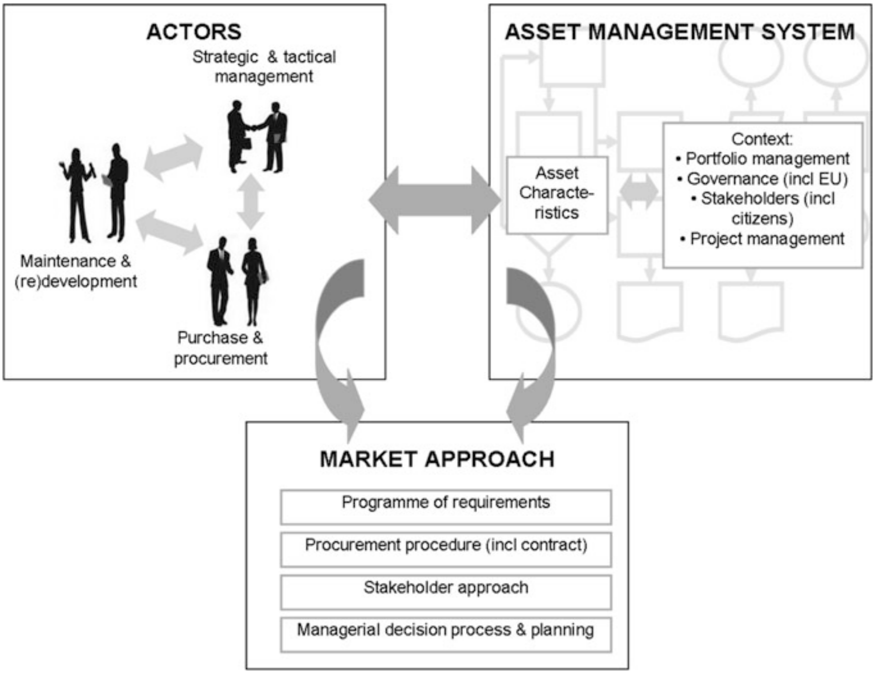


Fig. 57.3 Interaction between the actors at different levels in the organization, the asset management system and the market approach of an organization (based on [13])

situation, and interventions to the infrastructure throughout the life cycle of the asset), can be considered as the main challenge for effective asset management decision making. The three decision areas are intertwined, and it is this interrelationship that makes asset management decision making complex and dynamic. Volker [13] studied how principals acting in the public contracts deal with public procurement for design services in construction. Based on the findings in this research, the implications for asset management within an organization could be identified (see Fig. 57.3). Figure 57.3 shows the interaction between the actors at different levels in the organization, the asset management system, and the market approach of an organization. The strategic and tactical management level, the maintenance and (re)development department and the purchase and procurement department are the most important actors to make decisions regarding asset management. The asset management system consists of the assets and the context of the asset, which includes portfolio management, governance, stakeholders, and project management. Asset management interventions include the selection of projects, the method to apply to select projects, the identification of projects, and the choice to use intervention types with different levels of intensity, for example the renovation or rehabilitation of an asset [3, 12]. The market approach of organizations follows from the interaction between the actors and the asset management system and includes a program of requirements of the activities that need to be performed, a procurement procedure including a contract, the stakeholder approach, and a plan for the execution of the activities. In line with these concepts, Schraven et al. [12] conclude that asset management is effective when:

- Infrastructure objectives are used to evaluate the situation of infrastructure assets and the evaluation criteria are clearly derived from the objectives;
- Infrastructure interventions take the current and future situation of infrastructure assets into account and decision makers are able to cope with future uncertainties and changing requirements;
- Infrastructure interventions result in an infrastructure situation which is consistent with the infrastructure objectives;
- Infrastructure objectives are continually monitored and evaluated based on the infrastructure interventions applied and on unexpected changes of the infrastructure situation.

These results confirm the socio technical and interactive character of asset management systems and the need for alignment between the strategic, tactical, and operational level of decision making.

57.5 Conclusion

It is evident that a new set of skills is required for a successful implementation of innovative contracts; in the road sector as well as in all other infrastructure sectors experimenting with them [7]. A new approach that takes account of the dynamic nature of this environment and the contracts is required. Road authorities need to realize that drafting complete and totally self-enforcing contracts for integrated project delivery methods that combine design and construction tasks with long-term performance-based service agreements is just not possible. During the life time of such contracts, the environment of the contract changes, information about relevant external factors is updated, and even client needs and therefore contract goals will change. It is not viable to specify the legal consequences of every possible state of the world. The uncertainty and information-asymmetry that both players are confronted with when participating in the tendering of infrastructure projects, often lead to sub-optimal deals both for asset owners (often authorities) and service providers (contractors).

The impossibility of writing fully detailed contracts requires in fact a greater alignment between the strategic goals and the performance indicators used in contracts. A clear formulation of the boundaries of the playing field and the rules of the game needs to be agreed upon during the tender phase; such that if a renegotiation during realization is needed, the uncertainties will be solved according to the agreed priorities on a strategic level. A strategic market approach and quality assurance procedures are therefore needed that is performance based and based on interoperable databases [4]. In order to respond in a dynamic way to the trends in the construction sector, asset management strategies need to be integrated in the market approach and daily routine of the organization [13]. The right information will give the right foundation to enable a relationship between asset owner/manager and service provider that is based on trust. This will secure the quality levels not only from a technical point of view but also from a socio technical perspective. The first step toward success is to create understanding across the whole organization that these new procurement practices require a very different attitude and set of skills. Next to that, awareness is needed of public servants and contractors that they are dealing or about to deal with problems of a different and more dynamic nature. Gaming and other simulation techniques offer a great opportunity in this respect, as they allow practitioners to learn from their mistakes without carrying the costs of real life failures.

57.6 Future Research

The research of Altamirano [7] explored the interaction of the institutional choices (tendering procedure, granting criteria, intermediate control processes) with technical choices for the type of maintenance (heavy or light maintenance actions), the frequency and the intermediate quality of the road. The latter was simulated by a computer simulation model—Road Roles—with the maintenance plans of the contractor as an input and uncertainty/chance factors to simulate the uncertain future. All other aspects, in particular the institutional processes and agreements were played out in real-life by real people. The results provided an excellent illustration of the reciprocal impact of technical and institutional design and replacement (and maintenance) issues.

Further research and development of training tools that support public authorities in the successful implementation and management of innovative contracts and their dynamics are urgently required to prevent public values being compromised and society paying far too much as part of the learning process. Therefore we have included the idea of simulation and serious gaming in a new project called ‘Dynamic contracting’. The project is funded by the Next Generation Infrastructures Foundation and is set up as a collaborative project between the Faculty of Technology, Policy and Management, the Faculty of Electrical Engineering, Mathematics and Computer Science of Delft University of Technology and a research company called Almende. For this project we drafted a model that integrates the different phases for procuring asset management contracts: tender, execution and evaluation. Our aim is to connect the tender phase with the operational execution phase by supporting the planning and procurement decisions of both the asset manager and the service providers. In this way the strategic asset management decisions can be translated into the contracts. In order to increase flexibility of the road system, we will look at the design and maintenance activities from a network level. This means that a set of asset management activities will be connected to a group of service providers in the context of a network contract. By creating these kinds of conceptual networks, the dynamics and flexibility of the assets management should increase while the performance of the assets should stabilize on the stated performance level. Because asset management systems are socio technical by nature, we will make use of Agent Based Modeling Techniques [14, 15]. This makes it possible to model the system from the perspectives of the asset manager and service providers and take the life cycle perspective of assets into account. One of the major challenges will be to connect the requirements for the tender, contractual agreements, the monitoring system, performance and payment structure on a network level and make the model work in order to do simulations. Serious gaming will be part of this simulation in order to make actors learn to collaborate on a network level. We hope that in the end our system will support public agencies in their decision making processes for procuring asset management contracts and make a contribution toward filling the gap between the strategic, operational, and tactical level in asset management.

Acknowledgments This work was supported by the Next Generation Infrastructures Foundation (http://www.nextgenerationinfrastructures.eu/).

References

1. Dekker R, Scarf P (1998) On the impact of optimisation models in maintenance decision making: the state of the art. Reliab Eng Syst Safety 60(2):111–119
2. Herder PM, Wijnia YC (2012) A systems view on infrastructure asset management. In: van der Lei T, Herder P, Wijnia Y (eds) Asset management: the state of the art in Europe from a life cycle perspective. Springer, Dordrecht
3. Malano HM, Chien NV, Turral HN (1999) Asset management for irrigation and drainage infrastructure: principles and case study. Irrig Drainage Syst 13:109–129
4. Moon FL, Aktan AE, Furuta H, Dogaki M (2009) Governing issues and alternate resolutions for a highway transportation agency's transition to asset management. Struct Infrastruct Eng 5(1):25–39
5. Williamson OE (1998) Transaction cost economics: how it works: where it is headed. Economist 146(1):23–58
6. Bauer JM, Herder PM (2009) Designing socio-technical systems. In Gabbay DM et al (eds) Handbook of the philosophy of science: handbook philosophy of technology and engineering sciences, Elsevier, North-Holland, pp 601–631
7. Altamirano MA (2010) Innovative contracting practices in the road sector—cross-national lessons in dealing with opportunistic behaviour. Next generation infrastructures. PhD thesis, Delft University of Technology, Delft
8. Salmine A, Viinamäki O-P (2001) Market orientation in the finnish public sector. From public agency to privatised company. Helsinki, Ministry of Finance and University of Vasa
9. Cavana RY, Maani KE (2000) Systems thinking and modelling, understanding change and complexity. Prentice Hall, Auckland
10. Altamirano MA, Virtala P, Äijö J (2007) A qualitative analysis of periodic maintenance of roads. Understanding the mechanisms in pavement management. In: 25th international conference of the system dynamics society and the 50th anniversary celebration, Boston, USA
11. Demkes R et al (2007) Wegbeheer in nederland, feiten en ontwikkelingen. Rapportage fase 1 van het project anders organiseren van wegbeheer (AOWb) (Road management in the

Netherlands, facts and developments). Den Haag: Ministry of Traffic and Waterworks, RWS-SDG en DGP
12. Schraven D, Hartmann A, Dewulf G (2011) Effectiveness of infrastructure asset management: challenges for public agencies. Built Environ Proj Asset Manage 1(1):61–74
13. Volker L (2010) Deciding about design quality: value judgements and decision making in the selection of architects by public clients under European tendering regulations. Sidestone Press/Delft University of Technology, Leiden
14. Borshchev A, Filippov A (2007) From system dynamics and discrete event to practical agent based modeling: reasons, techniques, tools. In: Proceedings of the 22th international conference of the system dynamics society, Oxford, England, 25–29 July 2004
15. Schieritz N, Milling PM (2003) Modeling the forest or modeling the trees: a comparison of system dynamics and agent-based simulation. In: Proceedings of System Dynamics Conference, New York, USA

Chapter 58
The Dynamics of Outsourcing Maintenance of Civil Infrastructures in Performance-Based Contracts

R. Schoenmaker, J. A. de Bruijn and P. M. Herder

Abstract An important feature in managing civil infrastructures is the growing use of outsourcing in the delivery of maintenance. Evidence of this is shown in a strong increase in the application of performance-based contracts. The expectations of the principals are high: a smaller organization, better service, lower costs, more innovation, and more flexibility. But there are also risks connected to performance-based outsourcing of maintenance: the use of the wrong performance requirements, strategic behavior of the contractor, and a lack of knowledge and experience of the principal. The main question for the authorities that want to outsource their maintenance is: how do we achieve as much as possible of the expected advantages while limiting the possible disadvantages to a minimum? Based on case study and theoretical exploration this paper answers that question by investigating the strategies that the English Highways Agency and the Dutch Rijkswaterstaat use, when outsourcing the maintenance of their existing road infrastructures and what the effects of their strategies are. Lessons are drawn from the case studies and they are of most interest to other road authorities that consider, or already have chosen, outsourcing the delivery of maintenance as the way forward.

R. Schoenmaker
Department of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands
e-mail: r.schoenmaker@tudelft.nl

J. A. de Bruijn (✉) · P. M. Herder
Department of Technology, Policy and Management, Delft University of Technology, Delft, Netherlands
e-mail: j.a.debruijn@tudelft.nl

P. M. Herder
e-mail: p.m.herder@tudelft.nl

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_58,

58.1 Introduction

In recent years, outsourcing maintenance of civil infrastructures drew more and more attention of asset owners, asset managers, contractors, politicians and though more limited, of scholars. Traditionally public organizations, as asset owners, used their own staff for determining and delivering the required maintenance. The public organization used private parties to execute specific, prescribed tasks under direct supervision of public staff. Pressed by institutional reforms, public opinion, and limited financial means the public asset owners increasingly collaborate with private parties. Different forms of collaboration appear in varying combinations of design, build, maintain, operate and/or finance. The emphasis lies on using *performance-based contracts* because of the anticipated advantages that come with steering on outputs. In performance-based contracts, the focus lies on the results the principal wants to achieve and not on specifically mentioning all the required activities. In case of performance-based contracts for maintenance, the goal is to achieve those results, e.g., an available, safe, comfortable road, over a *longer period* for an *existing* infrastructure.

The general perception is that greater autonomy of the contractor and stronger links between performance and payment will lead to (cost) advantages for the principal. A critical note here is appropriate. The anticipated advantages of performance-based contracts are compared to the traditional way of contracting, in which the work was specified in much detail. Identifying and quantifying the advantages and determining a causal relation between the way of contracting and the advantages prove difficult [1].

There are not only advantages to performance-based contracting, but also disadvantages and risks, such as opportunistic contractor behavior, or difficulty in specifying the performance. The problem for the principals is, that with the introduction of contracts based on measuring performance, not only the advantages but also the disadvantages and risks will become manifest and the expected advantages will not be gained to full potential.

58.2 Research Question

In spite of the difficulties with assessing the real advantages and the associated risks, many asset owners turn to performance-based contracting. Not a great deal of research has been done on the effects and operation of performance-based maintenance contracts and what that means for the governance of those contracts [2–4]. The following question will be explored in this paper.

What tactics are used by public organizations when outsourcing maintenance of existing civil infrastructures to achieve as much as possible of the expected advantages while limiting the possible disadvantages to a minimum?

The next section discusses the concepts of road maintenance, outsourcing, and the combination of outsourcing and performance measurement.

58.3 Road Maintenance and Its Complicating Features

In this paper, maintenance is defined as: *The whole set of activities that are needed for keeping the required function(s) available at the agreed level of service.*

In addition to this definition it is essential that maintenance be executed in an efficient and effective way by minimizing risks, restoring and preventing defects and minimizing life cycle costs.

Also of importance is the cyclic nature of maintenance. Based on an existing model [5], a model of the cyclic maintenance process has been developed [6], see Fig. 58.1, that is used for analyzing the cases, see Sect. 58.5.

The performance requirements that have to be maintained are input to this model, followed by a cyclic process of measurement, analysis, work identification, planning, work preparation, and execution. This model is used to consistently describe the degree of outsourcing the maintenance.

Maintenance has a number of complicating features that make the use of performance measurement problematic:

1. The degree of complexity;
2. The degree of autonomy;
3. The degree of temporal mismatch;
4. The degree of innovation and dynamics;
5. The degree of unknown issues—*unknowns, unknowables*.

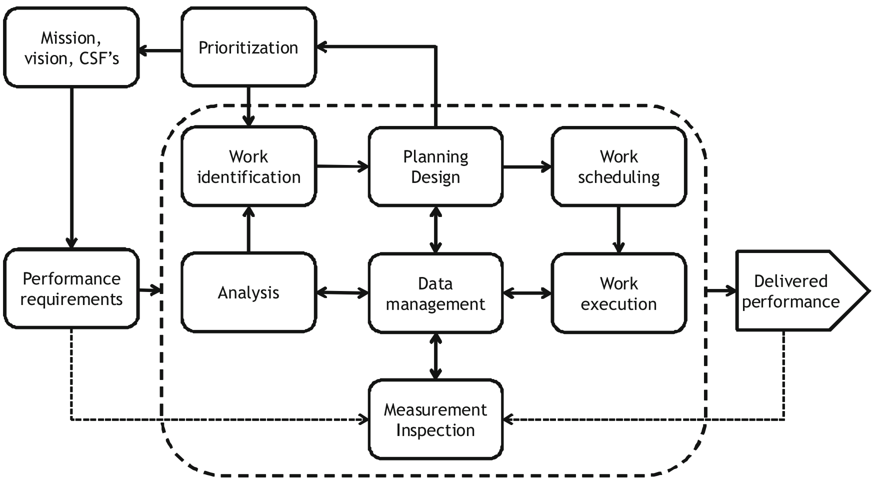


Fig. 58.1 Model of the cyclic maintenance process [6]

An example of complexity is the interaction of competing performance requirements, giving a choice of alternative measures, like a rapid response at higher costs or a more economic response with more hindrance to the customer. The contractor cannot always act autonomously because third parties are also at work in the same network or the contractor depends on third parties (permits, information, and supply) to be able to deliver. Temporal mismatch will occur when the effects of the maintenance activities will only show after completion of the contract. The dynamics are visible in the changing budgets of the principal, in the requirements (i.e., more emphasis on sustainability or economy) or in the developing technology, making possible news ways of working or even new levels of service. Examples of unknown issues are the costs of finding out the cause of defects and failures, the time and cost needed to repair damage after incidents or the necessity at the time of earlier scheduled maintenance [6].

Generally, the more parts of the maintenance process are outsourced and the more abstract (also multi-valued) the performance requirements are, the larger the influence of the complicating features. Because of these features, the principal will never achieve the full potential of the expected advantages of performance-based contracting.

58.4 Theoretical Perspectives

Three theoretical perspectives were adopted to explore the workings and difficulties of performance-based contracting in road maintenance.

Transaction cost economics: Transaction cost economics studies the most efficient governance structure for a transaction, such as contracts, alliances, hierarchies, etc., by looking at the characteristics of that transaction and assuming bounded rationality and the possibility of opportunistic behavior. Every governance structure has certain costs. Minimizing transactions costs leads to efficient organizational boundaries and a corresponding efficient governance structure [7, 8].

Agency theory: The agency theory studies the most efficient form of contract in a principal-agent relationship in different situations of goal congruence, information asymmetry, and a difference in risk-aversion between the principal and agent. The principal has a choice of different mitigating measures, each accompanied by a certain cost [9, 10].

Performance measurement: When studying performance measurement systems, focus is on designing and applying the process and structure of performance measurement in such a way that the positive effects are promoted and the inherent perverse effects are minimized. Performance measurement systems are never complete: not all aspects of reality will be covered by the measurement system [11, 12].

The theoretical perspectives result in an analytical framework that is used to study the practices in four cases. Elements of the framework are the characteristics of the transaction, of the relation, the incentives and of the process and structure of the performance management system.

58.5 Case Study and Case Protocol

The study comprises of a longitudinal study of several cases, by making new combinations from existing sources (contracts, evaluations, reports, audits, etc.). Interviews and expert meetings are used, aimed at validation, explanation, and illumination of the information from the written sources.

In order to be able to compare the strategies over a longer period, the research is longitudinal. The following four cases have been studied:

Netherlands:	Rijkswaterstaat (RWS)	Performance contracts	2004–2008;
Netherlands:	Rijkswaterstaat (RWS)	Pilot contract	2007–2010;
England:	Highways Agency (HA)	MA-TMC contract	2001–2008;
England:	Highways Agency (HA)	MAC contracts	2004–2010.

The theoretical framework is not usable for a one-on-one description of the reality. Describing and analyzing the 'contracted reality' is done using the research questions mentioned below. The relevant features of the theoretical perspectives will become visible using this operationalization of the analytical framework. The empirical findings are grouped under four themes:

1. *Theme: The object that is outsourced*
   (a) physical parts that are maintained and the degree of integration of the various technical disciplines
   (b) types of maintenance that is part of the outsourcing
2. *Theme: The division of roles*
   (a) division of roles in the maintenance process
3. *Theme: performance indicators and incentives*
   (a) types of indicators that are used
   (b) nature of the indicators that are used
   (c) payment mechanisms are used
   (d) other, more specific, incentives
4. *Theme: Governance*
   (a) governance of the contract

58.6 Case Description and Observations

Rijkswaterstaat is the executive agency of the ministry of Infrastructure and Environment in the Netherlands and manages three strategic networks: roads, waterways, and waterworks. The case study focuses on the road network of 3,250 km. The expenses for maintenance of that network were more than $ 1 billion in 2009. There were around 50 performance contracts that cover about 150 km of roads each. The pilot contract covered one area in the Netherlands and was used to gain experience with more integrated and more performance-based contracting.

Highways Agency is the executive agency of the Department for Transport in the United Kingdom and manages the strategic road network of 7,650 km in England. The budget for maintaining the network was $ 1.8 billion in 2009–2010. There were 24 MA-TMC contracts covering the whole of England. The MA-TMC contracts were gradually replaced by 14 (covering larger areas) MAC contracts.

The observations for each case are described in detail in Chaps. 5, 6, 7, 8 in Schoenmaker [6]. Below a summary of the observations is given.

Physical parts: Most contracts comprise nearly all the components of the infrastructure, e.g., gutters, verges, asphalt, bridges, viaducts, lining, rest areas, and signs. Both RWS and HA exclude technical installations for automated traffic management. Winter maintenance is part of the HA contracts, whereas RWS uses separate contracts.

Types of maintenance: RWS primarily includes routine maintenance plus a small amount of periodic maintenance in these contracts. The HA includes all maintenance activities up to a financial threshold.

Division of roles: In RWS cases the contractor has to comply with the performance requirements and no (change or quality) processes are described or prescribed. There is no process approach in the MA-TMC case of the HA. When describing the requirements, the activities are most significant. In the later MAC contracts the primary process is prescribed and the contractor has to submit a quality plan based on that process.

Type and nature of the indicators: RWS only uses product indicators (output) linked to the quality of the infrastructure, but not indicators that are linked to the overall goals of RWS. The HA uses input, throughput, and output indicators for product and non-product indicators, e.g., for cost and collaboration. The HA has put in a lot of effort in developing a performance management system that links the multi-valued goals and the activities needed to achieve those goals.

Payment mechanism: RWS uses a fixed payment for the routine maintenance as a whole and for each maintenance activity that is ordered separately. The HA uses three payment mechanisms: fixed price, hourly rates, and target prices.

Other incentives: Both principals initially used lane rental, but abandoned that principle due to safety issues. RWS used fines in the first case, but not in the second case. The HA does not use fines, but uses penalty points in case of shortcomings.

Those penalty points will lead to extra audits, at the cost of the contractor. In all cases there is the possibility of extending the contract.

The HA uses past performance as an incentive for good (present) performance, by using the present performance (by then past) as an important factor when selecting candidates for future tenders.

Governance: Both RWS and HA changed from direct supervision to governance based on the quality management system of the contractor. The RWS *prestatiebestek* is awarded on price only, the pilot contract on price/quality criteria. The HA uses that principle in both cases. Additionally the HA uses an instrument to measure process maturity of the contractor in the selection phase of the tender procedure. The HA emphasizes its partnering principles and pays a lot of attention to community building. RWS introduces the partnering principles in the pilot contract; the concept of community building is missing at RWS.

58.7 Case Study Results: Core Problems, Effects, and Comparison

The observations and findings briefly discussed in the previous section have been analyzed from each theoretical perspective, using the following set of questions: (1) What is the core problem as seen from the theoretical perspective? (2) What effects become manifest in the contracts studied? (3) What are differences in tactics between the cases in dealing with the core problem? and (4) What is the most efficient tactic according to each theoretical perspective?

Core problems: From each of the three theoretical perspectives, a core problem can be identified. (1) Transaction cost economics state that the presence of bounded rationality and (the chances of) opportunistic behavior lead to 'serious contracting difficulties' [7]. The bounded rationality of the actors makes the complexity of maintenance, the unknowables, and the uncertainty about (past and) future contingencies relevant. It is costly for both parties to the end the transaction untimely, because of the incurred transaction specific investments and the promise of future benefits in line with the longer duration of the transaction. Apart from the switching costs, the loss of reputation is also a driver for both parties to continue the transaction. It is beneficial to introduce safeguards by creating the right governance structure. Because the transaction lasts several years and is repeated over time, there is room for implementing a governance structure specifically for this transaction. The governance should be fitting to deal with the degree of uncertainty and complexity. That makes uncertainty, partly caused by complexity, the core problem for the transaction cost perspective. (2) The agency theory states several characteristics of the principal-agent relation that influence the choices for the most efficient balance between input and output steering [10]. The principal tries to achieve goal congruence by using output steering and performance management. Of the characteristics of the relation (like measurability of output/throughput, task programmability, degree of uncertainty) it is mainly the uncertainty about the behavior of the agent that causes contractual difficulties. Will the agent always act in the interest of the principal? And because in all cases there is evidence of incomplete monitoring, information asymmetry is seen as the core problem according to the agency theory. The goals of the principal and agent are not congruent and the principal is not fully informed about the behavior of the agent. There is a need for the principal to reduce the level of or the impact of the information asymmetry and the way of contracting needs to match that. (3) The cases show that strategic behavior is present in all of the cases. The consequences of the strategic behavior are nearly always sub optimization and inefficiency for the principal because the agent optimizes for his own benefit. The sub optimization and inefficiency result in higher costs of maintenance interventions either directly or becoming visible at later times, possibly after completion of the contract. A performance measurement system will never be perfect. In combination with the complicating features of maintenance, there will always be chance for strategic behavior making the possibility of perverse effects of performance measurement the third core problem. The multiple (and multi-valued) goals of the principal and the complicating features of maintenance and performance measurement make outsourcing problematic.

Effects: In all four cases, contracting difficulties are apparent, regardless of the approach used by the principal. This is proof of the persistent character of the perverse effects. The cases are rich in examples showing the effects that appear when the principal turns to outsourcing based on performance requirements, implementing an, until then not yet defined, approach to mitigate the contracting difficulties. In all cases governance was considered to have room for improvement by the respective national audit offices [13, 14].

A few examples: sub optimization because of the short term focus of the contractor based on the performance indicators, cherry picking when proposing maintenance interventions, lack of ambition when measures are easily met, insufficient contract management leading to invoices being paid without proper checking, etc.

Comparison of the approaches: RWS and HA have adopted clearly different approaches for dealing with the contractual difficulties. RWS opts for ex-ante completeness, for maintaining a certain distance to the contractor and for performance measurement that is solely used for measuring contract compliance. The HA opts for ex-post flexibility, interaction, and performance measurement that is, apart from the inevitable contract-compliance, mainly used for continual improvement, see Table 58.1.

Table 58.1 Approaches for dealing with contractual difficulties

Theory	Core problem	Rijkswaterstaat	Highways Agency
Transaction cost	Uncertainty	Ex-ante completeness	Ex-post flexibility
Agency theory	Info-asymmetry	Distance philosophy	Interaction, participation
Performance mgt.	Perverse effects	Focus on compliance	Focus on improvement

58.8 Most Efficient Approach Based on Theoretical Perspectives

In due course, both RWS and HA have moved to more abstract performance measures with, as a consequence, a growing influence of the complicating features of maintenance and performance measurement. This logically leads to a number of inevitabilities. According to the transaction cost economics the characteristics of the transaction (relation) for the HA (already to a great extent) and for RWS (to an increasing extent) demand a hybrid governance structure with ex-post negotiation. Decisions about future contingencies are taken during the course of the contract instead of at the outset. The essential components of a hybrid arrangement are:

- Classic contracting for performance measures that are less ambiguous and/or lead to activities with a predictable, repetitive character;
- Bilateral governance performance measures that are more ambiguous and/or lead to activities that are less predictable and a have a less repetitive character;
- Specific, agreed processes for dealing with the bilateral aspects of the governance.

The governance of such hybrid contracts thus consists of:

- Output steering linked with payment based on results;
- Input steering linked with payment based on effort;
- Input steering with programmed tasks and payment based on completed tasks;
- A division in the decision-making process, giving the principal influence on the choices the agent is making (proposing).

The performance measurement system calls for (ex-post) interaction and a partial decoupling of payment and performance (achievement). The interaction will lead to more trust of the contractor in the performance measurement (use and review). The partial decoupling of payment and performance (achievement) moderates the performance measurement and decreases the propensity towards strategic behavior.

58.9 Conclusion: Suggestions for Implementing the Most Adequate Approach

When taken to a more operational level, the most efficient approaches according to theory can be described as suggestions for implementing the most adequate approach:

1. *Adjust the governance to the degree of uncertainty:*

Only part of all performance requirements will lead to predictable maintenance activities that can be planned years ahead. For those requirements the entire maintenance process can be outsourced. When uncertainties begin to play a bigger role, a more intelligent process is needed. The essence of the intelligent process is a division in the maintenance process (see Fig. 58.1). The principal (i.e. asset manager) demands maintenance proposals from the contractor. The principal decides whether or not to implement those proposals. It is the obligations of the contractor to timely propose the maintenance activities in order to (enable the principal to) comply with all of the multi-valued goals.

2. *Provide nearness between actors and sharing of information:*

The intelligent process requires the sharing of information and the nearness of those involved. The sharing of information and the nearness of the actors have to be stimulated. First, by using performance measurement in a less coercive way than for contract compliance only. Second, by facilitating physical nearness of the actors through integrated teams and shared offices. Third, by having community building contribute to mutual understanding and trust and making that community *the* place for joint development and improvement of processes.

3. *Provide incentives for collaborative behavior:*

The uses of the shadow of the near future and the shadow of the far future have proven to be effective in the cases. The *shadow of the near future* is the outlook for more work during this contract. The *shadow of the far future* is the outlook for more work after this contract, based on the present performance during this contract. Good performance increases the chance of success 'there and later'. A third incentive is linked to the audits of the quality management system of the contractor. Positive results from audits can be used to award the contractor with a lighter audit regime. And vice versa.

4. *Put demands on the competencies of the outsourcing organizations:*

Outsourcing organizations have to be able to design and fine-tune the intelligent process and to manage the process at the interface with the contractor. This intelligent process requires a more complex management than when staying remote from the contractor, and it requires an *informed* en *knowledgeable* principal. The points mentioned under (2) should benefit becoming an *informed* principal. They have to reduce the information asymmetry between principal and contractor. As a consequence, the principal has to able to give a sensible meaning to the acquired information, hence the need for the principal to be *knowledgeable* and be able to take the right decisions.

References

1. Anastasopoulos PC, McCullouch BG, Gkritza K, Mannering FL, Sinha KC (2010) A cost savings analysis of performance-based contracts for highway maintenance operations. J Infrastruct Syst 16(4):251–263

2. Hyman WA (2009) Performance-based contracting for maintenance. National Cooperative Highway Research Program (USA), Washington DC (NY); Report No.: NCHRP Synthesis 389
3. Pakkala PA, de Jong WM, Äijö J (2007) International overview of innovative contracting practices for roads. Finnish Road Administration, Helsinki
4. Altamirano MA (2010) Innovative contracting practices in the road sector. Technische Universiteit Delft, Delft
5. Murthy DNP, Kobbacy KAH (2008) Complex system maintenance handbook., Springer series in reliability engineeringSpringer, London
6. Schoenmaker R (2011) The chosen way. Delft University of Technology, Delft
7. Williamson OE (1985) The economic institutions of capitalism. The Free Press, New York
8. Milgrom PR, Roberts J (1992) Economics, organization, and management. Prentice-Hall, Englewood Cliffs
9. Fama EF, Jensen MC (1983) Separation of ownership and control. J Law Econ 26(2):301–325
10. Eisenhardt KM (1989) Agency theory—An assessment and review. Acad Manag Rev 14(1):57–74
11. de Bruijn JA (2007) Managing performance in the public sector. Routledge, London
12. Meyer MW (2002) Rethinking performance measurement: beyond the balanced scorecard. Cambridge University Press, New York
13. National Audit Office (2009) Highways agency contracting for highways maintenance. The Stationery Office, London
14. Rekenkamer A (2009) Rapport bij het jaarverslag 2009 van Ministerie van Verkeer en Waterstaat (XII), in Tweede Kamer, vergaderjaar 2009–2010, vol 37(2), p 32

Chapter 59
Deterioration Prediction of Superstructure Elements of Community Buildings in Australia Using a Probabilistic Approach

H. Mohseni, S. Setunge, G. Zhang and P. Kalutara

Abstract Buildings are complex infrastructure assets and optimization of maintenance and rehabilitation actions require a well-considered asset management model. Community Buildings as one of the major investments of the local government in Australia have a large proportion of demand, expectation, and consideration among the local council's assets. The following paper introduces a building asset management (BAM) framework and a building element hierarchy which facilitates the building asset management and inspection strategies. Data gathering and preparation will be discussed followed by the calibration of a probabilistic deterioration prediction approach based on the Markov process. The Markov transition matrices have been derived for building elements based on the condition data sourced from local councils. The transition matrices for superstructure elements are presented, reviewed, and compared.

59.1 Introduction

To be certain about something is truly a rarity. We live and work in a complex world with uncertainties constantly surrounding us. Probability is a branch of mathematics devoted to quantifying the degree of uncertainty present in natural phenomena [1]. Probabilistic approaches are widely used in infrastructure management frameworks especially for roads and sewer systems.

H. Mohseni (✉) · S. Setunge · G. Zhang · P. Kalutara
RMIT University, Melbourne, VIC, Australia
e-mail: hessam.mohseni@rmit.edu.au

S. Setunge
e-mail: sujeeva.setunge@rmit.edu.au

G. Zhang
e-mail: zhangg@it.uts.edu.au

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_59,

'Asset Management' and 'Sustainability' are words frequently heard in local government and public works organizations throughout Australia. We have recently entered an era of extremely accurate measurement of assets, and have therefore gained a fine appreciation for the longstanding costs of sustainability [2].

The paper proposes an infrastructure asset management framework to be used for building assets. A three-level building element hierarchy is introduced for categorization of building components. Data preparation to be utilized in the non-linear optimization method for calibration of Markov transition matrices is discussed. Superstructure transient probabilities and expected condition curves are compared in the following sections.

The research work reported in this paper is part of a research project funded by the Australian Research Council and the data used for calibration of the model has been collected from six local councils of Victoria as partners of this project. The outcome of this research will be implemented at the partners' local authorities.

59.2 Framework

The conceptual framework of a Building Asset Management (BAM) practice is shown in Fig. 59.1. The structure of the framework has derived from not only a literature review, but also from the current asset management practices of this research project's partners and discussions with the experts. The continuous

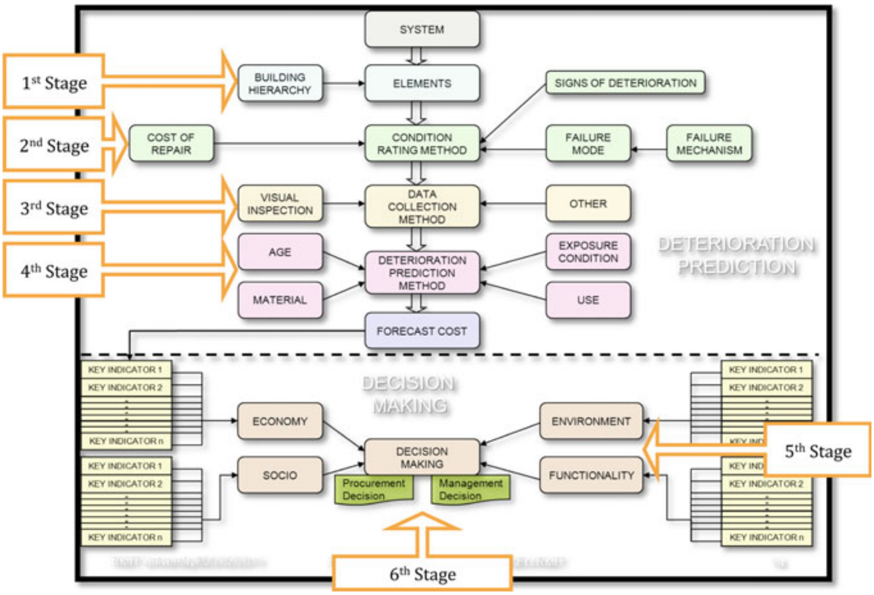


Fig. 59.1 Conceptual asset management framework

process which can be established by linking different stages and components will result in an appropriate decision making outcome in terms of not only building management, but also other infrastructure assets with some modifications [3].

59.3 Building Element Hierarchy

There are several different element hierarchies adopted by different organizations for their BAM practice. Consequently, a not only comprehensive, but also flexible element categorization has been used in this research to overcome the appropriate categorization issue. Therefore, a comprehensive building hierarchy based on the Building Component Guideline (BCG) introduced by NAMS [4] with a flexible categorization hierarchy has been proposed. The building element hierarchy divides the system into 10 main component groups (level one) which includes components such as Electrical Services, Exterior Work, External Fabric, Fire Services, Interior Finishes etc. The second level breaks each component group into several component types. Each component type will be divided further into several components in the third level of hierarchy. An example of this element hierarchy is depicted in Fig. 59.2.

The collapsible version of this three-level building element hierarchy, which consists of almost 400 components, introduces a convenient and flexible condition monitoring procedure and the future prediction feature in both core and advanced building management plans.

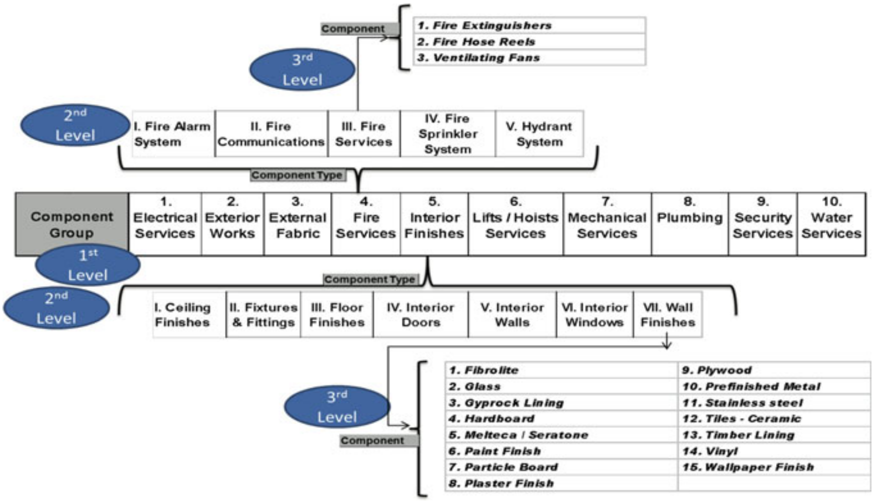


Fig. 59.2 Building element hierarchy

59.4 Deterioration Prediction Method

This section briefly describes the Markov chain theory and the non-linear optimization technique which has been adopted to derive the transition matrices for the Markov process.

According to Morcous [5], the Markov chain theory is still the most frequently used method in many statistical models. Simulating the degradation process is a complex but necessary task. This process is probabilistic in nature due to the uncertain environmental factors in the service life duration and the variability among each individual component. Hence it is desirable to simulate and predict this process in the framework of stochastic models. However, the validation and parameter identification of a stochastic model depends on the availability and format of data, coming either from controlled experiments or field tests [6]. According to Madanat, Markovian transition probabilities have been used extensively in the field of infrastructure management, to provide forecasts of facility conditions [7].

In deterioration modeling the attributes of a model randomly change over time. A Markov chain is a probability model, which has a finite-state, for describing a certain type of stochastic process that moves in a sequence of phases through discrete points in time according to fixed probabilities. The process is stochastic because it changes over time in an uncertain manner. In this chain the future states are dependent only on the present state and independent from the any state before the present states. Markov chain consists of transition matrix and initial distribution which can be introduced as a current state of an element. Transition matrix consist of a set of finite set of states S (1,1,3…n) and a propriety pi j to pass from state i to state j in one time step t. Time can be treated as either discrete (called Discrete-Time Markov Chain) or continuous (called Continuous-Time Markov Chain).

Before starting using the Markov Chain a condition rating method needs to be adopted. As an example a five-state condition rating can be proposed as conditions 1,2,…&5 where condition 1 represents the element in a brand new condition and 5 is the condition of an element with serious damage needing replacement.

Although the deterioration processes evolve over continuous time, for simplicity, discrete time steps could represent these processes (such as the time of the building inspection). Hence in this paper Discrete Time Markov Chain will be considered as a model for predicting the life cycle for building elements.

Discrete Time Markov Chain is a finite-state stochastic process in which the defining random variables are observed at discrete points in time. If an element is in state "i", there is a fixed probability, Pij of it going into state j after the next time step. Pij is called a "transition probability". The matrix P whose ijth entry is Pij is called the transition matrix. Transition matrix consist of a set of finite set of state S (1,1,3…n) and a propriety pi j to pass from state i to state j in one time step t. In Markov chain pi j should satisfy two following conditions.

$$p_{ij} \geq 0,$$

$$\sum_j p_{ij} \leq 1.$$

This mean if an element is in state i, there is a (P_{ii}) probability that this element will stay in state i, and ($1-P_{ii}$) will move to the next state j.

Present state at time t is i: $X_t = i$

Next state at time t + 1 is j: $X_{t+1} = j$

Conditional Probability Statement of Markovian Property:

$\Pr\{X_{t+1} = j \mid X_0 = k0, X_1 = k1, \ldots, X_t = i\} = \Pr\{X_{t+1} = j \mid X_t = i\}$

Discrete time means $t \in T = \{0, 1, 2, \ldots\}$

The probability of an element being in a given state at a given point in time can then be depicted by the set of curves shown in the empirical case study section.

An initial distribution 'v' is a single row matrix representing the number of elements in each state. In Markov chain after one time step the new distribution will be the result of multiplying initial distribution v by the transition matrix P

Distribution After 1 Step: vP

The distribution one step later, obtained by again multiplying by P, is given by $(vP)P = vP^2$.

Therefore distribution After 2 Steps = vP^2

Similarly, the distribution after n steps can be obtained by: vP^n

P^2 is the two-step transition matrix for the system. Similarly, P^3 is the three-step transition matrix, and P^n is the n-step transition matrix. This means that the ijth entry in P^n is the probability that the system will pass from state i to state j in n steps.

Tran [8] used the non-linear optimization technique based on the Bayesian Markov chain Monte Carlo Simulation method to calibrate the Markov model for stormwater pipe. In calibrating the Markov Models (i.e. estimating the transition probabilities), the Bayesian approach was used to estimate Pij from its posterior.

59.5 Deriving Transition Matrices from Condition Data: Non-Linear Optimization Technique

In Tran's study [8], the prior distribution $\pi_0(P)$ was arbitrarily chosen as a uniform distribution in the interval [0, 1], and since there was no available knowledge about the proper distribution of P_{ij}, the posterior distribution $\pi(P|Y)$ is proportional to the likelihood function $L(Y|P,M)$. Using the joint probability theory, the likelihood to observe set of pipe conditions (Y) can be expressed in Eq. (59.1), which was transformed into the logarithmic format for faster computing as in Eq. (59.2) [9].

$$L(Y|P,M) = \prod_{t=1}^{T}\prod_{i=1}^{3}\left(C_i^t\right)^{N_i^t}, \tag{59.1}$$

$$\log[L(Y|P,M)] = \prod_{t=1}^{T}\prod_{i=1}^{3} N_i^t \log\left(C_i^t\right), \tag{59.2}$$

where

t is the pipe age in years

T is the largest age found in the dataset

N_i^t is the number of pipes in condition I at year t

C_i^t is the probability in condition i at year t and can be computed by equation derived from Chapman-Kolmogorov formula [9].

If the initial condition state at year 0 is expressed by $C^0 = \left(C_1^0, C_2^0, C_3^0, C_4^0\right)$ of being in one of the four condition states at year t can be computed using the equation below.

$$\left(C_1^t, C_2^t, C_3^t, C_4^t\right) = \left(C_1^0, C_2^0, C_3^0, C_4^0\right) \cdot \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ 0 & P_{22} & P_{23} & P_{24} \\ 0 & 0 & P_{33} & P_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{59.3}$$

where;

C_i^t is the probability being in the condition state i at year t

C_i^0 is the probability being in the condition state i at year 0

$$\sum_{i=1}^{4} C_i^t = 1, t = 0, 1, 2, 3, 4 \rightleftharpoons$$

59.6 Data Collection and Implementation

This section presents data preparation and modifications and the assumptions made to customize the data for application of the non-linear optimization technique. Modeling assumptions and the implementation of the non-linear optimization technique in accordance with calibration of the Markov transition probabilities and acquiring the prediction scheme are also presented.

The data used in this study, provided by one of the research project's partner councils, were sourced from two condition data sets collected in 2007 and 2009 of around 270 building assets. Analysis of elements of a major hierarchical sub-component (superstructure) is presented in this paper; where superstructure

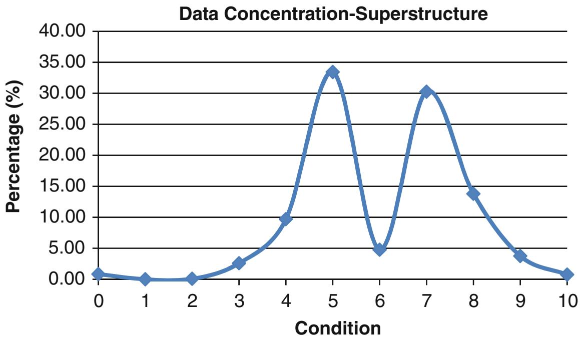


Fig. 59.3 Data concentration

comprises ceiling, external doors, external wall, internal doors, internal screens, internal wall, roof, stairs, upper floors, veranda post, and windows.

Some modifications have been made to the input data including not only elimination of outliers' and removal of inconsistent and incomplete data, but also omitting the elements on which maintenance and/or rehabilitations identified to be carried. Hence, the data left for 203 building assets have been used for this paper.

According to Howard et al. [10], the age of elements can be calculated by the equation V, therefore ages have been calculated accordingly.

$$\text{Useful life} = \text{Age} + \text{Remaining life}. \tag{59.4}$$

The age consistencies were checked and verified via two different data sets. As a result, data of 167 buildings have been shown to be reliable for initial analysis.

The condition monitoring plan with which the data were collected has considered 11 condition states (from 0 to 10) where; 0 is considered as new condition and 10 represents very bad or a failing condition. Figure 59.3 shows the percentage of data concentrated in different condition stages for superstructure subcomponent. The figure VII has been generated from 1196 superstructure elements inspected (Fig. 59.4).

Two major observations can be made from the data concentration graphs. Firstly, the data sets have a very small number of elements in good condition stages and secondly, the major concentration of the data is in the mid-condition stages with the maximum number of data is on the fifth condition stage. Since the buildings inspected aged from 8 to 82 years, a question can be raised regarding the accuracy of the data. One possible hypothesis discussed is that the uncertainty of the inspectors and/or condition rating manual/guideline used in rating elements led to majority of data being concentrated in the mid-condition rates.

In order to use the data for the calibration of the transition matrices and the Markov chain generation for acquiring transient probabilities, 0–10 condition states have been consolidated into 1–5 condition states [3].

Superstructure/Internal Walls -					
Cond.	1	2	3	4	5
1	0.51	0.11	0.37	0.01	0.00
2	0.00	0.54	0.21	0.23	0.02
3	0.00	0.00	0.91	0.08	0.00
4	0.00	0.00	0.00	0.99	0.01
5	0.00	0.00	0.00	0.00	1.00

Superstructure/External Walls -					
Cond.	1	2	3	4	5
1	0.24	0.44	0.31	0.00	0.01
2	0.00	0.54	0.46	0.00	0.00
3	0.00	0.00	0.85	0.15	0.00
4	0.00	0.00	0.00	0.97	0.03
5	0.00	0.00	0.00	0.00	1.00

Superstructure/Roof - Transition					
Cond.	1	2	3	4	5
1	0.43	0.15	0.29	0.12	0.01
2	0.00	0.45	0.54	0.00	0.01
3	0.00	0.00	0.93	0.07	0.00
4	0.00	0.00	0.00	1.00	0.00
5	0.00	0.00	0.00	0.00	1.00

Superstructure/Ceiling - Transition					
Cond.	1	2	3	4	5
1	0.44	0.20	0.21	0.14	0.01
2	0.00	0.63	0.29	0.08	0.00
3	0.00	0.00	0.97	0.03	0.00
4	0.00	0.00	0.00	0.97	0.03
5	0.00	0.00	0.00	0.00	1.00

Superstructure/Stairs - Transition					
Cond.	1	2	3	4	5
1	0.54	0.00	0.25	0.21	0.00
2	0.00	0.49	0.35	0.15	0.01
3	0.00	0.00	0.98	0.01	0.01
4	0.00	0.00	0.00	0.99	0.01
5	0.00	0.00	0.00	0.00	1.00

Superstructure/Internal Doors -					
Cond.	1	2	3	4	5
1	0.31	0.40	0.23	0.05	0.00
2	0.00	0.50	0.50	0.00	0.00
3	0.00	0.00	0.90	0.10	0.01
4	0.00	0.00	0.00	1.00	0.00
5	0.00	0.00	0.00	0.00	1.00

Superstructure/External Doors -					
Cond.	1	2	3	4	5
1	0.60	0.10	0.27	0.02	0.01
2	0.00	0.51	0.00	0.49	0.00
3	0.00	0.00	0.93	0.06	0.01
4	0.00	0.00	0.00	0.99	0.01
5	0.00	0.00	0.00	0.00	1.00

Superstructure/Windows -					
Cond.	1	2	3	4	5
1	0.37	0.23	0.30	0.10	0.00
2	0.00	0.59	0.29	0.11	0.01
3	0.00	0.00	0.94	0.06	0.01
4	0.00	0.00	0.00	1.00	0.00
5	0.00	0.00	0.00	0.00	1.00

Superstructure/Upper Floors -					
Cond.	1	2	3	4	5
1	0.63	0.00	0.36	0.00	0.01
2	0.00	0.73	0.27	0.00	0.00
3	0.00	0.00	0.93	0.07	0.00
4	0.00	0.00	0.00	0.31	0.69
5	0.00	0.00	0.00	0.00	1.00

Superstructure/Veranda Post -					
Cond.	1	2	3	4	5
1	0.37	0.06	0.50	0.07	0.00
2	0.00	0.77	0.23	0.00	0.00
3	0.00	0.00	0.94	0.06	0.00
4	0.00	0.00	0.00	1.00	0.00
5	0.00	0.00	0.00	0.00	1.00

Superstructure - Transition mtrix					
Cond.	1	2	3	4	5
1	0.84	0.00	0.15	0.01	0.00
2	0.00	0.54	0.25	0.17	0.04
3	0.00	0.00	0.84	0.12	0.04
4	0.00	0.00	0.00	0.85	0.15
5	0.00	0.00	0.00	0.00	1.00

Fig. 59.4 Superstructure transition matrices

59.7 Analysis of Results

The process of calibration and derivation of the transition matrices was in accordance with the Sect. 59.5. Figure 59.5 illustrates transition matrices of all individual elements of the superstructure and also the transition matrix of the superstructure as a sub-component. Figure 59.6 depicts the transient probabilities

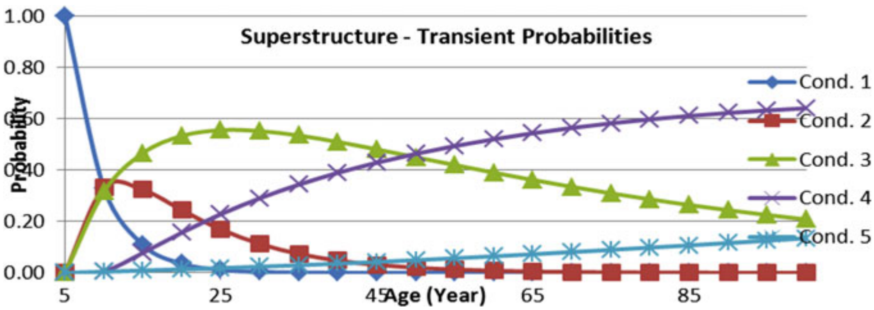


Fig. 59.5 Superstructure transient probabilities

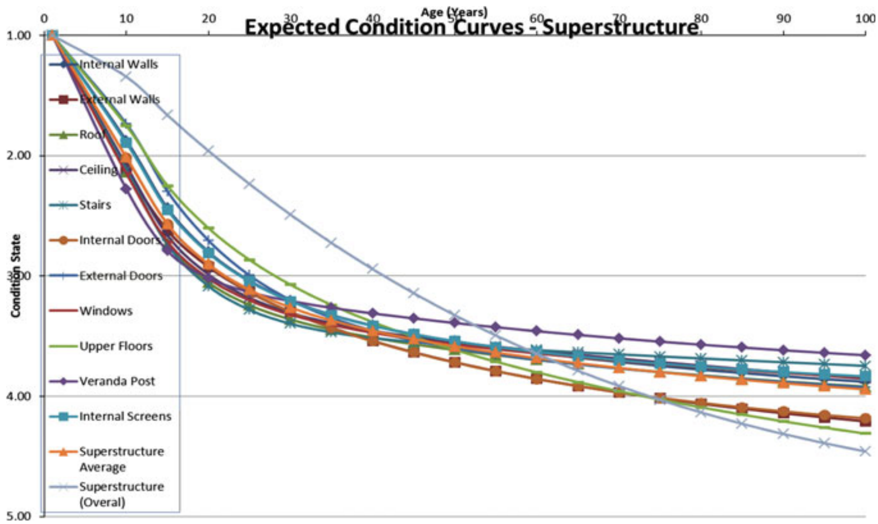


Fig. 59.6 Superstructure expected condition curves comparison

of the superstructure and the Fig. 59.6 shows a comparison of the expected condition curves of each element and the superstructure sub-components.

The transient probabilities of other elements can be generated by use of their transition matrix in accordance with the method described in Sect. 59.4.

It can be seen that in some elements such as "Windows" and "Internal Doors" the transition matrices do not show any transition from the condition 4 to the condition 5. As mentioned in Sect. 59.6, this is because of the lack of inspected elements in condition 5 in the model input data. Although Fig. 59.6 is not showing a probabilistic prediction of elements, could be a good guide for predicting the future states of elements and also understanding the challenges of acquiring reliable condition data sets.

59.8 Conclusions

The paper presented the derivation of a Markov chain based deterioration curves for superstructure of community buildings. The probabilistic deterioration curves as well as the expected condition curves derived using the data have been presented. The expected condition curves indicate that the variation between deterioration curves of different super structure elements is quite low. However, the probabilistic estimates shown in Figs. 59.5 and 59.6 do present a significant variation between different elements. In addition, the expected condition curves show the sensitivity of the outcome between the case of using individual components data and using the data related to all elements of one sub-component (superstructure). The proposed approach for deriving deterioration curves using Markov theory can be highly sensitive to the condition data as shown. A reliable and representative data set is essential for derivation and calibration of the models.

References

1. Miller GK (2006) Probability—Modeling and applications to random processes. Wiley, Texas
2. Hain A (2011) The 'as designed & as constructed' process. Public Works Engineering, June 22–July 3, 2011
3. Mohseni H, Setunge S, Zhang G, Edirisinghe R, Wakefield R (2011) Deterioration prediction of community buildings in Australia. Paper presented at: ICOMS asset management conference, Gold Coast, Australia, 19–20 May 2011
4. Building Condition & Performance Assessment Guidelines (2009) I—NAMS
5. Morcous G, Rivard H, Hanna AM (2002) Case-based reasoning system for modeling infrastructure deterioration. J Comput Civil Eng 16(2):104–114
6. Zhang Y, Augenbroe G (2005) Uncertainty analysis in using Markov chain model to predict roof life cycle performance. Paper presented at: International Conference on Durability of Building Materials and Components, Lyon, France, 17–20 April 2005
7. Madanat S, Wan Ibrahim WH (1995) Poisson regression models of infrastructure transition probabilities. J Transp Eng 121:267–272
8. Tran HD (2007) Investigation of deterioration models for storm water pipe systems. PhD thesis, Victoria University
9. Estimated Sewer Management Systems. J Eng, Construct Architectural Manage
10. Howard J, Dixon J, Comrie J (2006) Determining remaining useful life. Australian infrastructure financial management guidelines, 5 edn

Chapter 60
iFactory Cloud Service Platform Based on IMS Tools and Servo-lution

A. Kao, J. Lee, E. R. Lapira, S. Yang, Y. Huang and N. Yen

Abstract Cloud computing has forced companies to a paradigm shift on how services are managed and delivered, and this includes how prognostics and health management (PHM) applications are implemented. Currently, companies are constrained in developing fleet-wide or distributed monitoring systems due to limitations in computing resources, difficulty to achieve scaled application deployment, among others. This paper addresses these issues through the iFactory Cloud Service Platform which integrates a cloud infrastructure (Servo-lution) and PHM analytics. A tool condition monitoring (TCM) module has been developed as a service application and has been successfully demonstrated on a horizontal machining tool test-bed.

A. Kao (✉) · N. Yen
Institute for Information Industry, Taipei, Taiwan
e-mail: hakao@iii.org.tw

N. Yen
e-mail: nelson@iii.org.tw

J. Lee · E. R. Lapira · S. Yang · Y. Huang
Center for Intelligent Maintenance Systems, University of Cincinnati, Cincinnati, USA
e-mail: jay.lee@uc.edu

E. R. Lapira
e-mail: edzel.lapira@uc.edu

S. Yang
e-mail: shanhuyang@gmail.com

Y. Huang
e-mail: huangyx@uc.edu

J. Lee et al. (eds.), *Engineering Asset Management 2011*,
Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-1-4471-4993-4_60,

60.1 Introduction

Prognostics and health management (PHM) is an emerging engineering discipline that evaluates the reliability of a system within its life-cycle in order to detect beforehand any upcoming failures and reduce risks. PHM connects failure mechanisms to system life-cycle management [1]. Knowing failure of certain equipment in advance and preventing it would save a significant amount of time and money while increasing the overall reliability and safety of both products and operations. Currently, more and more research effort has been shifted toward PHM, which focuses more on incipient failure detection, current health assessment and remaining useful in life prediction.

However, while more and more factories are now emphasizing on PHM, the difficulties are exposed in PHM applications to factory level since it is not easy to reconfigure deployed PHM application parameters in different factories. Besides, some PHM tasks that use highly sampled data and complex algorithms may require high computing resources, which are expensive for medium or small scale industries to own, operate and maintain an IT infrastructure to support a multi-site and distributed monitoring system. Therefore, since the technology of cloud computing is becoming mature, it can aid machinery industry to conduct PHM with higher feasibility. The cloud computing is opening a new era for the industries on how services are managed and delivered, and this also includes how Prognostics and Health Management (PHM) applications are implemented.

This paper addresses these issues through the iFactory Cloud Service Platform which integrates a cloud infrastructure (Servo-lution) and PHM analytics (Watchdog Agent® Tools). A Tool Condition Monitoring (TCM) system has been developed as a practical service application and successfully demonstrates the cloud-based TCM and PHM on a horizontal machining tool test-bed.

60.2 Smarter Machinery Maintenance

In the past decade, many research initiatives started in an effort to provide proactive maintenance strategies to manufacturing industries. The goal is to achieve highly consistent product quality, process performance, and fewer breakdowns. The proposed integrated scalable predictive maintenance system platform based on cloud computing enables a machine to implement self-prognostics and self-maintenance capabilities via standard protocol and cloud-based infrastructure. Critical sub-systems and components like spindle, feed axes and many others are identified. Prognostics models are developed as well. External addon sensors and controller signals are used for degradation monitoring and generating health information. A machine level health is generated by combination of the individual health of the critical sub-systems and their components. *MTConnect*, a middleware standard, is used to equip the machine to share the health information alongside

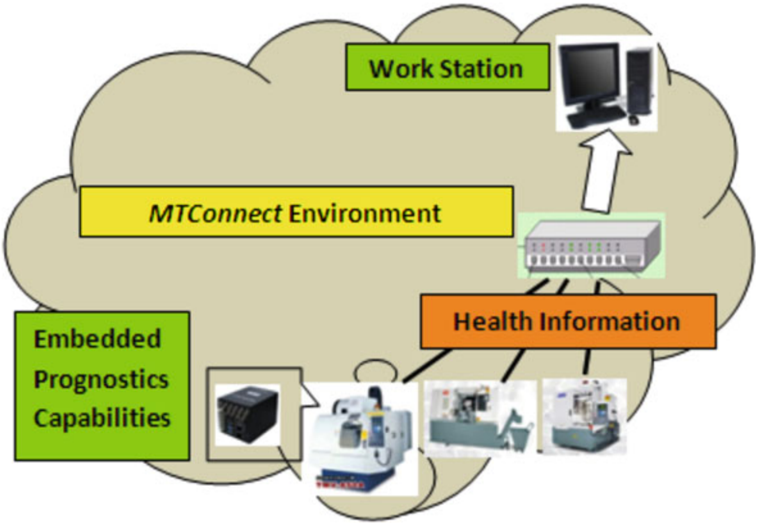


Fig. 60.1 Current smart machine with *MTConnect* environment (*source* annual reviews in control 35, 2011)

other important machine data to higher level systems for further processing with the usage of universal XML based standard. *MTConnect* does not provide the capability to analyze the data. It is used only for communication of the information for further analysis and decision support.

Figure 60.1 shows a current system where the machine with embedded prognostics tools has the capability to share the health information with the higher level systems using *MTConnect* middleware. In this environment, machine has no local capability of performing detailed analysis on the sensory data collected and will use external workstations for analysis.

60.3 The Infrastructure of Cloud Manufacturing: Servo-lution

To improve the feasibility of PHM, enabling machine tools to communicate with a cloud service platform is somehow a critical technology need to be solved first. The communication between machine tools and the service platform may include collecting and transferring data from machining process, and connecting to the platform for analytical service. Thus, Servo-lution is developed for machine tool builder and machinery industry. The Servo-lution for integrating machine tool industry includes four dimensions: machine tool, ServBox, applications, and platform. The first vision of Servo-lution is to create *one* intelligent platform, which is not only in local factory but also connected to satellite factories,

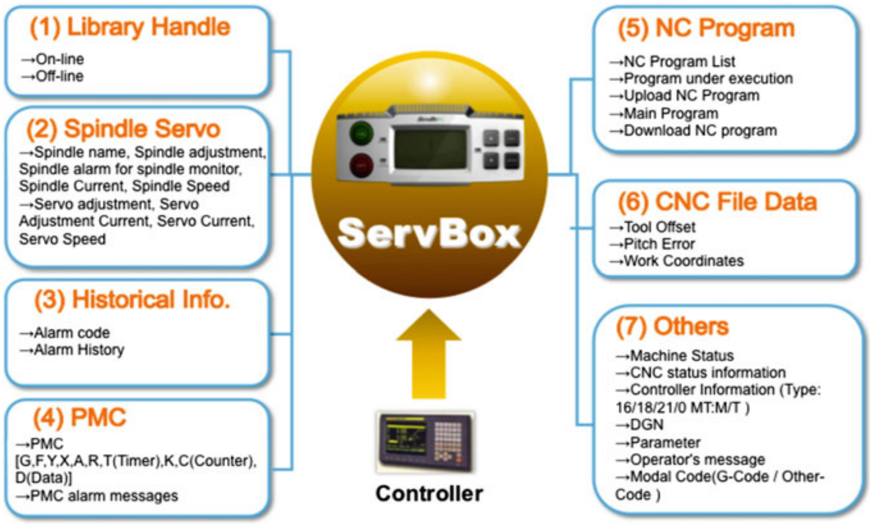


Fig. 60.2 Parameters can be extracted from controller through ServBox

therefore, customers can easily know the amount of the products produced. Second, there will be several services being brought into *two* endpoints. One endpoint is machine tool, and the other is manufacturers. Finally, three benefits including machine tool information transparency, sustainable preservation of information and real-time reaction to the problems are delivered to the customers.

There are three different parts in Servo-lution including ServBox, ServAgent, and ServCloud. The following sections will introduce each part in detail.

60.3.1 Manufacturing Service Channel: ServBox

ServBox is a service channel that provides synchronized communication between the CNC machine tool and the remote service platform, so called intelligent Service Platform (iSP), from the machine tool builder. With both ServBox and service platforms including ServAgent and ServCloud, machine tool builders or machinery industry could deliver innovative services and machine network via remote cloud computing service to achieve optimal productivity and enhance the partnership with customers. Through ServBox, we can change the traditional model of service communication, and achieve an active and fast two-way communication with service platform.

Nowadays, ServBox can extract more than 4,000 parameters from Fanuc i-series controllers, including 6 main categories: historical information, CNC File data, NC Program, Spindle servo, library handle. The feature categories are listed in Fig. 60.2. When machine is in a healthy status, it records only the general parameter and machining log from machine tools. When some problem occurs, it can start detailed status monitoring by using seamless switching technique to enable a cloud-based PHM analysis. The parameter set recorded can be set with regard to the request of industries.

60.3.2 In-Factory Resource Management: ServAgent

ServAgent is a platform for increase productivity and collaborative manufacturing with machine tool builder. When there are more and more machine tools in a factory, ServAgent can grasp the machine tools' real-time utilization, monitoring operation and detecting machine performance. In addition, it can connect to machine tools only when needed and avoid data-leakage problem. Whether production department of MTB or satellite factory and machinery company, ServAgent can add machine intelligence together with ServBox. A win–win situation is created for MTB and industrial company by integrating the solution and production line based on the mechanism of production management.

ServAgent leverages the machining information extracted from machine tools to unveil the machining data that in the past is concealed in the mysterious manufacturing process. It collects productivity information from production lines and forms a machine service network to help customers manage their production line in real-time and more intelligent, and finally the productivity and utilization can be improved. While the domain-specific knowledge of machining process is accumulating, it can also support customer to find out the weakness in the current manufacturing process, which means the quality of their manufacturing is also increased. Besides, the knowledge collected can used to be an expert or reference and provided to customers. Before the abnormal condition occurs, take the actions to avoid the possible damages and stabilize the production quality level. Finally, customers can manage the overall productivity and efficiency of their equipment in the factory. It offers customers the flexibility and knowledge toward their equipment and factory.

60.3.3 Intelligent Service Platform: ServCloud

The architecture of ServCloud is shown in Fig. 60.3. On the left side is machine tool builder, and on the right is the machinery company. If every machine tool equips a ServBox, all of the machining information will be collected by ServBox, and then send to ServAgent. All information will be kept in ServAgent and not revealed to the ServCloud, so there is no data leakage problem. Moreover, ServAgent can provide the advantage of data transparency during manufacturing process. The only time that machinery companies want to share their data might be the time when there are machine failures and throughput drops. In this kind of

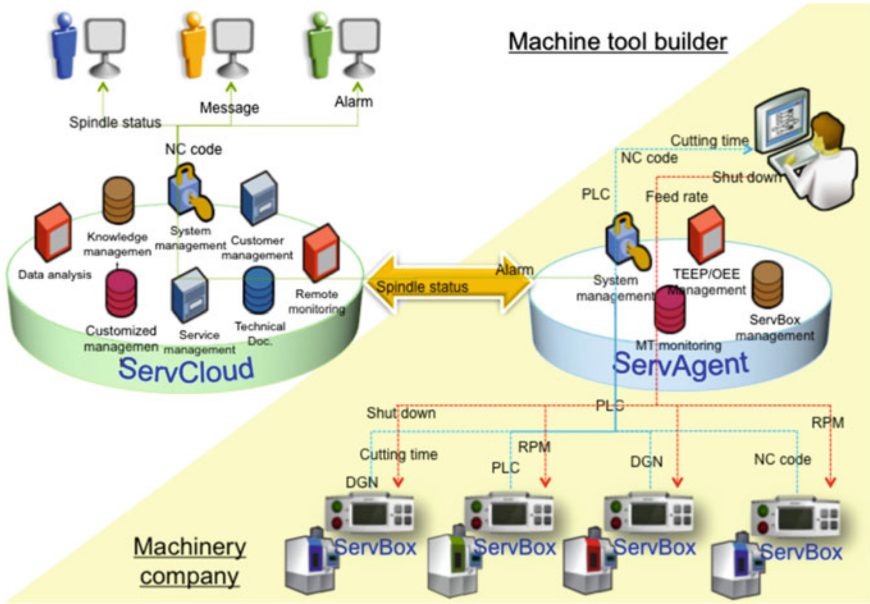


Fig. 60.3 ServCloud architecture

situation, they will actively transfer these data to ServCloud platform to let machine tool builder examine and recover machine failures immediately.

60.4 Application Case: Tool Condition Monitoring

Modern machine tool is a complex mechanical system that is integrated with a sophisticated controller. Despite technological advancements that the machine tool industry has gone through, it still experiences downtime due to component breakdown and wear. Thus, this is a good candidate for prognostics and health management or PHM by tracking failure precursors (features) and using analytic tools (such as those found in the Watchdog Agent® Toolbox developed by the Center for Intelligent Maintenance Systems) to estimate the condition of the component.

A major source of machine tool downtime that can be avoided is excessive wearing and breakage of cutting tools during machining operations. There has been significant scientific effort put into advancing tool condition monitoring or TCM and exhaustive surveys can be found. In terms of condition estimation, neural networks dominate the literature although other methods have also been applied such as fuzzy theory and hidden Markov models (HMMs). The popularity of TCM techniques highlight its importance in the machine tool industry as it can

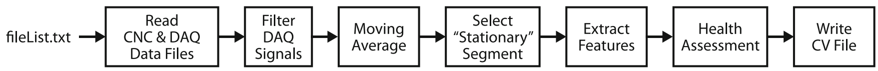


Fig. 60.4 TCM process flowchart

enhance productivity, produce and maintain better quality of machined parts and eventually lessen expenditures associated with automated manufacturing systems, usually associated with high volume production.

From the machine tool, tool wear sensitive signals such as spindle power and vibration are collected and digitized by the ServBox. Although they are not directly related to tool degradation, selected controller signals are also recorded in order to properly segment within the sensor signals. Both data streams are then sent to the ServCloud for temporary storage. A segmenting module is then initiated to remove leading and trailing samples that are not significant to the actual cutting operation. The remaining data segments are then stored for processing by the TCM module which produces a health state estimate for a given test data. The results are then written to a file that is used to update a web client graphical user interface (GUI) for visualization purposes. The TCM process flowchart that was used to generate the service application for use within ServCloud is shown in Fig. 60.4. When the TCM module is triggered, the program automatically searches for the file "fileList.txt". The file paths indicated in this ASCII file are then located and the associated files are parsed. Figure 60.5a shows the two sensor signals plots: vibration (top) and power (bottom) which was recorded after each cut. Figure 60.5b illustrates a combination of different controller signals.

The resulting signal undergoes a series of processes wherein features are extracted from a stable portion of the signal. A stable portion is defined as the duration of the data wherein the cutting tool is actually engaged onto the workpiece. Figure 60.6 shows the power data has been undergone a moving average process, after which, the stable part of the segment was identified using a k-means method. Intuitively, more power is expended when the tool is in contact with the part being machined, and the points marked with red diamonds in Fig. 60.6a indicate the points associated with the stable portion. Time location of the stable portion is used to isolate the equivalent segment in the vibration data. Features are then computed from the stable portion from both the vibration and power signals. Summary statistics such as average, standard deviation, minimum and maximum values are derived. Figure 60.6b shows three extremely correlated and relevant features (power mean & max and vibration mean) for an experiment wherein the tool made 12 passes or cuts. Observe the almost monotonic increase in the feature values which suggests that these features are sensitive inputs for tool health assessment.

The selected features are then fed to a health assessment technique which uses a Euclidean distance metric, $\boldsymbol{d}_{st}$.

$$d_{st} = \sqrt{(x_s - x_t)(x_s - x_t)'}, \tag{60.1}$$

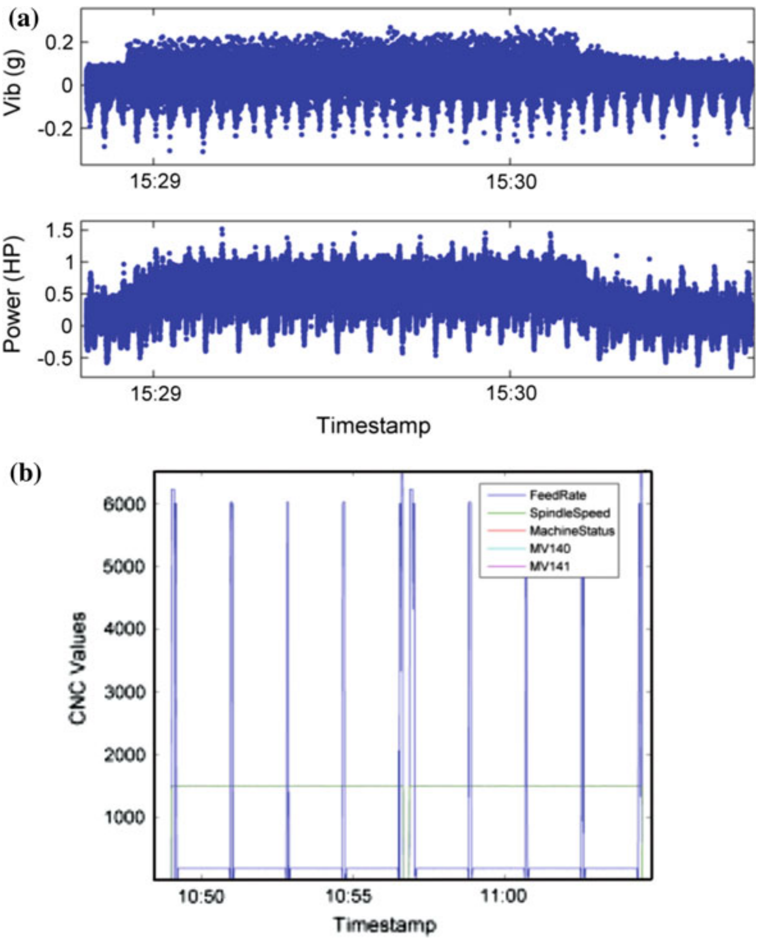


Fig. 60.5 Sensor signals (**a**) and controller signals (**b**)

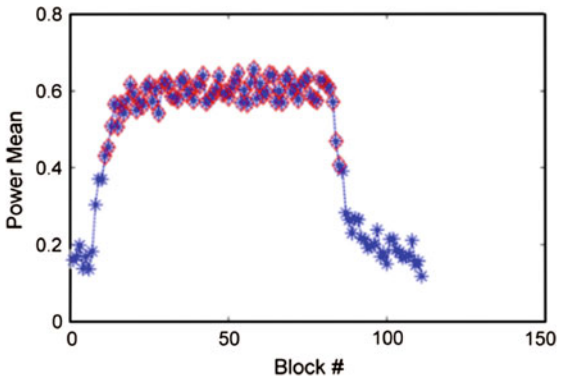
where x_s is the feature vector of the reference cut (first pass) and x_t is the feature vector of the test cut (current pass). The current distance metric is unbounded and a normalization technique is applied to generate the confidence value or ***CV*** that is constrained between [0,1]:

$$CV = (d_{thres} - d_{st})/d_{thres}, \tag{60.2}$$

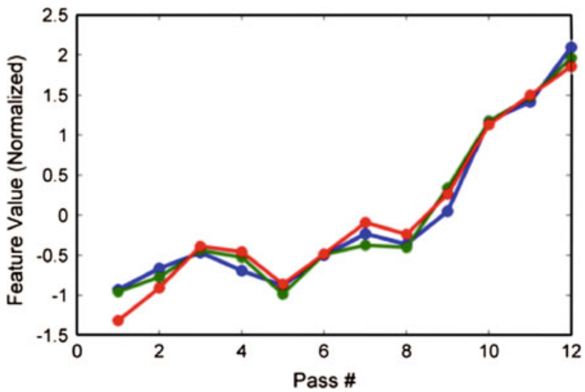
where d_{thres} was empirically found from previous cases when the machining conditions are exactly similar.

The health assessment results with normalized CV are then shown in the top plot of Fig. 60.7. Observe how the CV value starts out high (1 means the healthy status) and as the cutting tool is continuously used, the degradation manifests as an almost monotonic decrease in the health value. Eventually, the tool gets replaced

Fig. 60.6 Feature selection. **a** Power mean after moving average. **b** Select TCM features

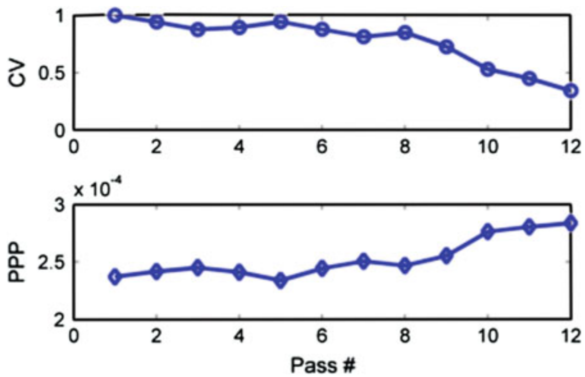


(a) Power mean after moving average



(b) Select TCM features

Fig. 60.7 CV results and average consumed power per pass



when CV reaches a value just below 0.5. The CV value when the tools gets replaced is relatively similar the cutting tests performed under similar machining conditions and parameters. The bottom plot of Fig. 60.7 depicts the average power

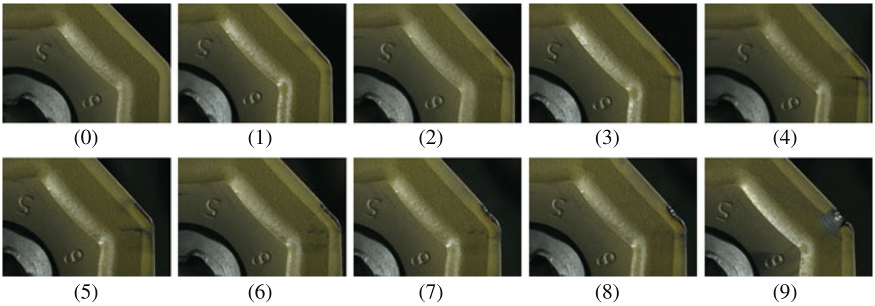


Fig. 60.8 The images for microscope view of the degradation process of a tool insert. The number under each image is the number of each cut

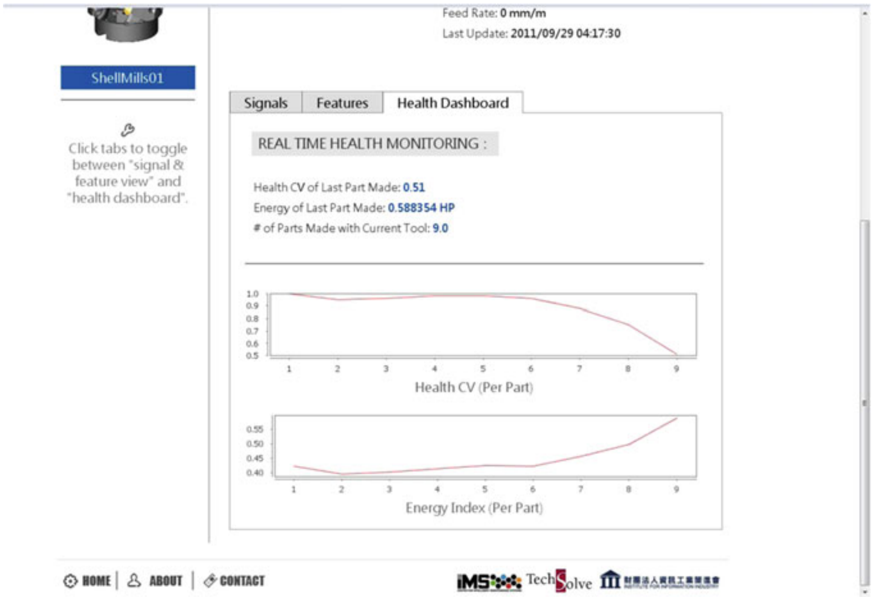


Fig. 60.9 A screenshot of the cloud-based prognostics for the tool insert (the horizontal axis indexes are the order of cuts)

consumed when a part is machined. Intuitively, the power consumed increases with tool wear.

Figure 60.8 shows a degradation progress of a tool insert. Each image is taken after a cut. And Fig. 60.9 is a screenshot from the cloud-based tool monitoring system. As shown in the figures, the CV decreases according to the degradation progress of the tool insert. The prognostic model is effective and can clearly deliver the prognostic information to the user.

60.5 Conclusions

The cloud-based machine service platform including Servo-lution and IMS tools embraces four advantages: (1) Incremental scalability: access to additional resources on-demand in response to increased application (e.g. computing) loads; (2) Agility: resources can be distributed among the users; (3) Reliability and fault-tolerance: built-in redundancies enable high levels of availability and reliability for continuous monitoring; (4) Utility-based: users only pay for the services they use.

This study implemented a cloud-based machine health monitoring infrastructure, and developed a case study on tool condition healthy monitoring. IMS Watchdog tool are adopted to conduct PHM on milling tools on milling machine and showed a rapid PHM task deployment. In the past, manufacturers cannot precisely identify the degradation of tools after each machining process except they use instruments to measure tools. However, measurements are labor-intensive and take time. Through the proposed cloud-based machine health monitoring service, manufacturers can easily understand the health condition of tools via iFactory user interface, thus the cost can be reduced and the workpiece quality would be improved since tools health condition now can be managed.

Acknowledgments This study is a collaborative research program of Center for Intelligent Maintenance Systems, Institute for Information Industry, and TechSolve Inc.

Reference

1. Uckun S, Goebel K, Lucas PJF (2008) Standardizing research methods for prognostics. In: Proceedings of the international conference on prognostics and health management, Denver, USA