



The northernmost University of Technology in Scandinavia
World-class research and education

Proceedings of

The 2nd international workshop
and congress on eMaintenance

Dec 12-14
Luleå, Sweden

eMaintenance

Trends in technologies & methodologies
challenges, possibilities and applications

2012



LULEÅ
UNIVERSITY
OF TECHNOLOGY



eMaintenance 2012

The 2nd International Workshop & Congress on eMaintenance

Kulturens hus, Luleå, Sweden
12-14 December 2012

Supported by:



*Operation and Maintenance Engineering
Luleå University of Technology*

Printed by Universitetstryckeriet, Luleå 2012

ISBN 978-91-7439-539-6

Luleå 2012

www.ltu.se

Editorial

eMaintenance has reached a high degree of attention within the industry. It has emerged from the integration of disciplines, such as; operation & maintenance engineering, software engineering, system engineering, information management, business management, together which it make up the fundamental of this domain. eMaintenance as an approach and a concept has been developed to support maintenance decision making process in enterprises through utilization and fusion of enabling eTechnologies; including eMonitoring, eDiagnostics, and ePrognostics. eMaintenance is being exploited in many industrial domains, with an objective to align the maintenance process with the business and operational processes to achieve organizational objectives. The benefits of eMaintenance are reflected in organisations' overall performance, efficiency, and effectiveness.

After the successful conduct of the 1st International Workshop and Congress on eMaintenance during June, 2010, time has arrived to conduct the 2nd International Workshop and Congress during 12-14 Dec 2012, at Lulea. We have received excellent support from both industry and academia in terms of number of technical papers and number of participants. The eMaintenance Workshop and Congress is planned to provide a regular platform every alternate years to initiate discussion amongst various partners to provide directions for effective utilization of ICT, besides the new and emerging technologies. eMaintenance solutions are the fusion and integration of various emerging technologies and methodologies. Thus, a number of challenges and issues related to wide domain of disciplines related to eMaintenance are included and considered during the congress.

The purpose and theme of the eMaintenance congress is to provide a timely review of research efforts on the topic covering both theoretical and applied research that will contribute towards understanding the strategic role of eMaintenance in asset management and performance of operation and maintenance of complex systems. The presentations and papers included in this proceeding cover most of the areas related to the main themes of the congress, as mentioned below:

- ❖ eMaintenance for Data warehousing & data mining
- ❖ eMaintenance for Embedded system technologies
- ❖ eMaintenance for Remaining Useful Life
- ❖ eMaintenance-Operation Performance Monitoring
- ❖ eMaintenance for Performance Based Logistic
- ❖ eMaintenance for Enterprise Resource Planning
- ❖ eMaintenance for Integrated Logistics Support
- ❖ eMaintenance-Prognostic Health Management
- ❖ eMaintenance Concepts and frameworks
- ❖ eMaintenance - Diagnosis and prognosis
- ❖ eMaintenance for Data quality
- ❖ eMaintenance for Information logistics
- ❖ eMaintenance for Decision support
- ❖ eMaintenance for Data and service fusion
- ❖ eMaintenance-Maintenance optimization
- ❖ eMaintenance - Life Cycle Management

Some selected papers from the proceedings will be published in the International Journals.

We thank all the authors for their papers and the reviewers for their reviewing support. We would also like to thank all the members of the International Advisory Committee, the Programme and Organizing committees, besides the colleagues and researchers of the Division for their active support. Our special thanks to Phillip Tretten and Karina Wandt, for their willing support and hard work.

We hope the participants and the readers will find this proceeding useful and informative.



Uday Kumar



Ramin Karim



Aditya Parida

Division of Operation and Maintenance Engineering
Luleå University of Technology, Luleå, SWEDEN

Conference Chair

Professor Uday Kumar
Luleå University of Technology

International Scientific Committee

Professor Benoit Iung
Nancy University, France

Professor Anita Mirijamdotter
Linneaus University, Växjö, Sweden

Professor Andrew K. S. Jardine
University of Toronto, Toronto, Canada

Professor Ajit K. Verma
IIT Bombay, India

Professor Per-Olof Larsson Kråk
The Swedish Transport Administration, Sweden

Professor Jay Lee
University of Cincinnati, USA

Professor Raj BKN Rao
COMADEM International

Professor Andrew Ball
Huddersfield University, UK

Professor Tore Markeset
Stavanger University, Norway

Professor Gao Jinji
Beijing University of Chemical Technology, china

Professor Stephan Heyns
University of Pretoria, South Africa

Professor Rakesh Mishra
Huddersfield University, UK

Professor Mira Kajko-Mattsson
Royal Institute of Technology, Sweden

Associate Professor Peter Söderholm
The Swedish Transport Administration, Sweden

Dr. Aitor Arniaz
Tekniker, Spain

Dr. Erkki Jantunen
VTT, Finland

Dr. Jan Holmström
Aalto University, Finland

Dr. Olov Candell
Saab Technologies

Dr. Seppo Saari
KTUAS, Finland

Dr. Gopinath Chattopadhyay
CQ University, Australia

Dr. Peter Funk
University of Mälardalen, Sweden

Dr. David Mba
School of Engineering Cranfield University, UK

Dr. Marco Macchi
Politecnico di Milano, Italy

National Committee

Dr. Olov Candell
Saab Technologies

Mr. Tony Järleström
Progressum

Mr. Gunnar Axheim
Vattenfall

Mr. Anders Kitok
LKAB

Mr. Sten Askmyr
LKAB

Mr. Björn Eklund
Trafikverket

Mr. Jan-Erik Nygård
BnearIT

Mr. Anders OE Johansson
ProcessIT Innovations

Mr. Anders Jonsson
ProcessIT Innovations

Organising Committee

Associate Professor Ramin Karim
Luleå University of Technology

Associate Professor Aditya Parida
Luleå University of Technology

Associate Professor Diego Galar
Luleå University of Technology

Assistant Professor Alireza Ahmadi
Luleå University of Technology

Dr. Phillip Tretten
Luleå University of Technology

Mr. Anders Jonsson
ProcessIT Innovations

PhD candidate Karina Wandt
Luleå University of Technology

PhD candidate Mustafa Aljumaili
Luleå University of Technology

Editorial Committee

Professor Uday Kumar
Luleå University of Technology

Associate Professor Ramin Karim
Luleå University of Technology

Associate Professor Aditya Parida
Luleå University of Technology

Dr. Phillip Tretten
Luleå University of Technology

**MINIMUM
LIFE CYCLE
COST
OR
MAXIMUM
LIFE CYCLE
CONFUSION?**



**INFORMATION
MANAGEMENT
BY SAAB**

www.saabgroup.com



SAAB



The hands-on experience

eMaintenance LAB

Contents

- **The Use of Laser Based Trolley for Railway Switch and Crossing Inspection** 1
Marius Rusu, Clive Roberts and Stephen Kent
- **Data to Decision through Contextual Presentation of Railway Infrastructure Performance** 7
Christer Stenström, Karina Wandt and Aditya Parida
- **Prognostic and Health Management of Wheel Condition: Integration of Wheel Defect Detection and Wheel Profile Monitoring Data** 15
Matthias Asplund, Stephen M. Famurewa and Matti Rantatalo
- **Improvement of Configuration Management in Railway Signalling System** 23
Amparo Morant, Phillip Tretten, Ramin Karim and Diego Galar
- **A handheld Maintenance Workstation: Information fusion in the aircraft – ground systems gap** 33
Torbjörn Fransson and Olov Candell
- **Simulation as Support for PBL Contract Design** 37
Olle Wijk and Patric Andersson
- **Applying an Integrated Systems Security Engineering Approach within Aviation eMaintenance** 43
Tim Tinney and Olov Candell
- **Augmented Reality as a New Level for Maintenance Efficiency** 49
Rúben Oliveira, José Torres Farinha, Inácio Fonseca, Luís Andrade Ferreira and Fátima Armas
- **Current and Prospective Information and Communication Technologies for the E-maintenance Applications** 57
Jaime Campos
- **Case-Based Reasoning Supports Fault Diagnosis Using Sensor Information** 63
Ning Xiong, Tomas Olsson and Peter Funk
- **Framework for Design and Performance Evaluation of Effective Condition Monitoring and Maintenance Management Systems** 69
Reza Farahani and Michael G. Lipsett
- **Usability Aspects of eMaintenance Solutions** 77
Karina Wandt, Phillip Tretten and Ramin Karim
- **Reliability Predictions based on Breakdown Data with Respect to Aging Conditions** 85
Jakob Krause and Klaus Kabitzsch
- **Condition monitoring of oil and gas subsea electrical equipment - Case study Åsgard subsea compression project** 89
Hermanto Ang, Tore Markeset and Tor Ole Bang-Steinsvik
- **A Study of Wireless Vibration Sensors for Monitoring Bearing Faults** 95
Mustafa Aliwan, Fengshou Gu and Andrew Ball
- **Re-Design of AC Excitation Busduct based on Infrared (IR) Thermography: Condition-Based Monitoring (CBM) data analysis** 101
Subramani Gopinath and Jamin Abdul Razak
- **Machine Health Monitoring: An Integrated Maintenance Approach** 107
Mahantesh Nadakatti, Sarbjeet Singh, Aditya Parida and Uday Kumar
- **Numerical Simulations of Effects of Faults in a Vertical Axis Wind Turbine's Performance** 113
Atif Shahzad, Taimoor Asim, Kyoo-Seon Park, Suman Pradhan and Rakesh Mishra
- **Computational Fluid Dynamics based Performance Optimisation of Vertical Axis Marine Current Turbines** 119
Kyoo-Seon Park, Taimoor Asim, Rakesh Mishra, Atif Shahzad and Gilbert Mirangwanda
- **Application of Particle Swarm Optimization Approach in the Inflationary Inventory Model** 125

under Stochastic Conditions

Seyed Mostafa Orand, Abolfazl Mirzazadeh and Farzaneh Ahmadzadeh

- **Multi-objective Reliability Allocation Problem** 131
Jalal Safari and Farzaneh Ahmadzadeh
- **On the Integration of Wear Model into Dynamic Analysis for Rolling Element Bearing** 137
Idriss El-Thalji and Erkki Jantunen
- **Control Charts for Assessment of Linear Asset Condition using both Temporal and Spatial Information** 145
Bjarne Bergquist and Peter Söderholm
- **Information Logistics for Continuous Dependability Improvement** 153
Per Norrbin and Peter Söderholm
- **Holistic Maintenance Information with Multi-layers of Contractors** 161
Mikael Palo, Isabelle Lindsund and Per-Olof Larsson-Krâik
- **E-Monitoring of Operation and Maintenance Task under Difficult Visual Work Environment** 167
Ravdeep Kour and Ramin Karim
- **Assessment of Reliability Data for Traction Frequency Converters using IEEE std 762 – A Study at Swedish Railway** 171
Yasser Mahmood, Ramin Karim and Mustafa Aljumaili
- **New Idea to Estimate the Reliability and Failure Probability with quantitative Methods for Critical Applications** 175
Ossmane Krini and Josef Börçsök
- **Modified Physics of Failure approach for Reliability prediction of Electronic Components** 181
Adithya Thaduri, Ajit K. Verma, Gopika Vinod, Rajesh Gopinath and Uday Kumar
- **eMaintenance Ontologies and Data Production** 191
Mustafa Aljumaili, Karina Wandt and Ramin Karim
- **Information Security in eMaintenance – a study of SCADA Security** 197
Yamur K. Al-Douri, Mustafa Aljumaili and Ramin Karim
- **e-Monitoring for Haul Road Maintenance in Mining Applications** 205
Stephan Heyns, Harry Ngwangwa, Theo Heyns and Stephanus van der Westhuizen
- **CMMS Benchmarking Development in Mining industries** 211
Yonas Lemma, Stephen M. Famurewa, Diego Galar, Håkan Schunnesson and Jonas Fjellner
- **Urban Road Transports E-maintenance supported in HMM** 219
António Simões, José Torres Farinha, Inácio Fonseca and Fernando Maciel Barbosa
- **A New Approach for the Diagnosis of Maintenance State** 229
Hugo Raposo, José Torres Farinha, Inácio Fonseca and Luís Andrade Ferreira
- **Lean versus Kanban - An Actual Paradigm?** 235
Pedro Dias, José Torres Farinha, Inácio Fonseca and Luís Andrade Ferreira

The use of laser based trolley for railway switch and crossing inspection

Marius Rusu
BCRRE
University of Birmingham
B15 2TT, Birmingham, UK
+44 121 414 2626
mfr028@bham.ac.uk

Clive Roberts
BCRRE
University of Birmingham
B15 2TT, Birmingham, UK
+44 121 414 4306
c.roberts.20@bham.ac.uk

Stephen Kent
BCRRE
University of Birmingham
B15 2TT, Birmingham, UK
+44 121 414 5063
KentSY@bham.ac.uk

ABSTRACT

Railway turnouts are used to guide trains from one track to another and they are usually formed of a switch and a crossing (S&C). They pose both economic and safety issues since they need to be maintained, which is costly, and if not in a good state they can cause fatal railway accidents. The purpose of this research is to improve the inspection process for S&C, which is accomplished by following a two step process that comprises of: (1) identification of areas where research is needed and (2) research and development of a laser based trolley for the inspection of railway switches and crossings. The first step was carried out systematically and is described in the second chapter of this paper. With the output of the first step as well as the help of Network Rail, it has been concluded that in Great Britain there is a need for a lightweight device that can carry out inspection for specific derailment hazards on switches and crossings as well as advise the welders on how they should be repaired. This research proposes a new approach for this issue by considering a lightweight laser based trolley system. The system is based on the ability to scan and record 2D slices of the parts that need to be inspected and calculate different wear indicators and derailment hazards in accordance with the Network Rail inspection standards. The work and findings of this research will also feed into Work Package 3 of the European project AUTOMAIN, Augmented Usage of Track by Optimisation of Maintenance, Allocation and Inspection of railway Networks, <http://www.automain.eu/>.

Keywords

railway, switch, crossing, automatic inspection, laser, profile measurement

1. INTRODUCTION

Railway turnouts make up a considerable part of any railway network and their purpose is to safely guide trains from one track to another. A turnout is usually formed of a switch and a crossing (S&C).

As stated by the UIC in a report concerning switches and crossings, their maintenance takes up a considerable amount of money from the total maintenance budget: "Switches and crossings (S&C) are a serious cost driver and take up about 25%

to 30% of the total maintenance and renewal budget each year"[1]. Because of certain features, switches and crossings are exposed to higher forces than plain line and, therefore, they degrade more rapidly than plain line. The maintenance of switches and crossings is very important to maintaining a healthy infrastructure. The inspection and repair procedures are vital to maintaining a good level of safety and keeping life cycle cost low.

This research focuses on improving the inspection methods for railway switches and crossings and it does this by following a two step process:

- 1) identification of areas where research is needed, and;
- 2) research and development of a laser based trolley for the inspection of railway switches and crossings.

The first step of this research was carried out as a sequence of three tasks. The first task was gathering information about inspection tasks which must be carried out on switches and crossings. This was done by consulting relevant inspection standards, categorising and recording in an Excel spreadsheet file all the inspection tasks that apply to switches and crossings. The second task was the identification of all available solutions that have the potential of carrying out any of the S&C inspection tasks. The third task was the identification of technology gaps where research would improve the inspection process.

After the completion of the first step, the findings led to the consideration of a laser based trolley that would improve the inspection of S&C. Several other factors endorse the research of such a system.

Safety Considerations

S&C must be well maintained, otherwise they can be a threat to the railway; one of the most serious consequences is a train being derailed. The derailments at London Waterloo on 11th September and 24th October 2006 [2] as well as that which occurred near Glasgow Exhibition Centre station on 3rd September 2007 [3] have a common element. Two of the major causes were the poor grinding work that had been carried out before the derailment, as well as the inability of the workers to correctly assess the switches in accordance with the Network Rail standard NR/L2/TRK/0053. More information on this standard can be found in the third chapter of this paper.

Network Rail Endorsement

Network Rail have identified that in Great Britain there is a lack of understanding of how weld and grind repairs should be carried out and that a combination of poor repairs together with incorrect profile measurements leads to poorly maintained S&C.

The challenge

It has therefore been concluded that in Great Britain there is a need for a lightweight device that can:

- 1) advise the maintenance team on how the S&C should be repaired by means of welding and grinding;
- 2) carry out specific profile inspections on switches and crossings, and;
- 3) do it efficiently.

Such a device would have a considerable impact on the railway by improving the way in which the repairs are being carried out and therefore improving the life cycle cost, and also by partially automating the inspection procedure on S&C, thereby considerably reducing human error and decreasing the probability of a derailment occurring.

Available Technologies

Currently, there are two technologies which are considered as possible solutions for S&C inspections. Although these off-the-shelf solutions have considerable potential to carry out the necessary tasks, for several reasons it was concluded that an alternative system must be researched. The applicability and limitations of these systems are discussed in the forth chapter.

Research hypothesis

This research proposes a new approach to this problem, and it does so by considering a lightweight laser based trolley which is able to both guide welders on how to repair the S&C and carry out an inspection for derailment hazards with little or no preparation required.

Paper contents

The second chapter describes the process of identifying areas where research is needed. The third chapter describes the issue in more detail. The forth chapter offers an insight into the available technologies and their limitations. The fifth chapter describes the proposed solution. The sixth chapter states a few conclusions.

2. IDENTIFICATION OF RESEARCH AREAS

The first step of this research was the identification of areas where research is needed. This was carried out as a sequence of three tasks, as depicted in Figure 1.

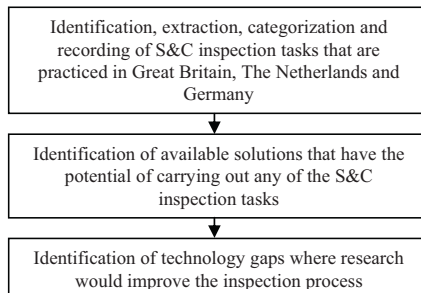


Figure 1. The process of identification areas of research

The identification of inspection tasks was done by consulting the railway inspection standards from Great Britain, The Netherlands and Germany. The reason for inspecting the standards from three different countries was to have a more generic overview of the inspection tasks, rather than considering just British standards. Each identified inspection task was categorised and recorded in an Excel spreadsheet file along with the following remarks: the country where the inspection task is practised, the inspection standard and the chapter and/or page number that sets out the inspection task, the frequency of inspection, and a remark about the current way of carrying out the inspection (e.g. using a manual gauge). An example showing the excel spreadsheet file that contained the gathered inspection tasks can be seen in Figure 2.

After the inspection of all relevant standards, the research continued with the investigation of (potential) available solutions.

Part of S&C	Countries	Inspection task	UK inspection standard that defines it	NL inspection standard that defines it	DE inspection standard that defines it	Average frequency of inspection	Current method used to carry out the inspecti
Stock rail inspections	UK, NL, DE	Side wear - measurement of plain rail side wear using NR4 gauge (NR); only in closure panel and transition zones (NL); measured in mm (DE)	NR/L2/TRK/0053, Chapter 8.4, Chapter 13.1	UIC Working Group Switches & Crossings, page 28	RiF S21 - Oberbau inspizieren, page 108	once a month to once a year	manually with a gauge
	NL, DE	Top wear - measurement of plain line top wear in closure panel and transition zones (NL); measured in mm (DE)		UIC Working Group Switches & Crossings, page 29	RiF S21 - Oberbau inspizieren, page 109	once a month to once a year	manually with a gauge

Figure 2. Example of S&C inspection tasks

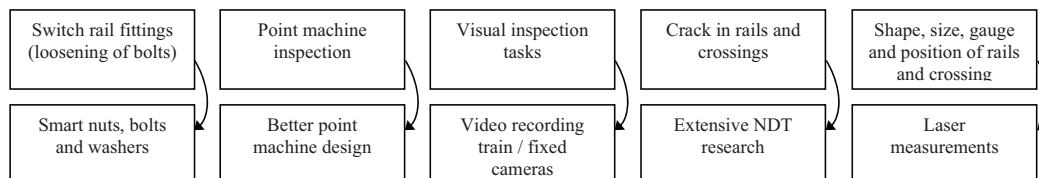


Figure 3. Areas of research

Therefore, the market for inspection technologies for switches and crossings was fully researched and all relevant available technologies were mapped in the Excel file in a separate column.

The difference between automatic inspection and condition monitoring must be noted here. These systems have in common a general purpose, which is to know the condition of the infrastructure. The approaches are, however, very different.

Condition monitoring allows infrastructure managers to know whether or not there is a fault, what kind of fault it is and, even better, predict when a fault will occur in the future. These features are known as: fault detection, fault diagnosis and fault prediction. In order to achieve this, a selection of sensors is made which can measure the parameters which give the most useful information. The information can be then processed using algorithms and fault detection, diagnosis or/and prediction can be achieved.

Automatic inspection is used to replace manual inspection. Manual inspections are set out by the inspection standards which provide a safe and accurate way to determine the condition of the infrastructure (if carried out accordingly). When automatic inspection is able to replicate manual inspection, then the condition of the infrastructure is known.

In conclusion, automatic inspection techniques can replace manual inspection, while condition monitoring techniques can help to know the condition of the infrastructure, but cannot replace manual inspection. This is because the output of the condition monitoring techniques is not as accurate and reliable as required by the inspection standards. For this reason, the condition monitoring solutions were recorded, but they are not candidates for automatic inspection solutions.

The last step was the identification of technology gaps where the current technology is not able to provide a solution. This step revealed prospective areas of research where new innovation may improve the way inspection is carried out. Five broad potential areas of research were identified and each inspection task was then categorised into one or two of these areas. Figure 3 shows these five areas of research. The laser based trolley approach is part of the "Shape, size, gauge and position of rails and crossing" area.

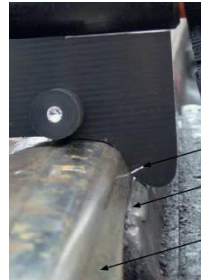
3. THE ISSUE

3.1 Network Rail's inspection standards

Network Rail, the infrastructure manager for most of Great Britain's railways, uses a set of two inspection standards to assess the condition of S&C. The standards set out procedures for assessing the profile of the S&C as follows:

- NR/L2/TRK/0053 – Inspection and repair to reduce the risk of derailment at switches
- NR/L2/TRK/1054 – Inspection, maintenance and repair procedures for cast, welded and fabricated crossings in the track

In Network Rail, the profile of S&C is assessed by the use of manual gauges which are placed on the inspected part(s).



60° line marked on wheel profile gauge. Gauge should not make contact with rail below this marker
 Switch blade angle visibly flatter than sidewear on stock rail
 Left-hand stock rail showing sidewear extending below top of switch blade (step 9)

source: http://www.raib.gov.uk/cms_resources.cfm?file=071218_R442007_Waterloo.pdf

Figure 4. NR P8 manual gauge applied on a switch

Figure 4 shows such a manual gauge, the shape of a P8 train wheel profile, which is used to assess if a certain derailment hazard is likely to occur. The manual gauge is placed on top of the two rails. The main rail on the left hand side of the picture is called the stock rail and the smaller rail on the right hand side is called the switch rail. The inspection result will be a pass only if the edge of manual gauge, which is below the white marker, is not in contact with any of the two rails. In this example the result is a fail. The readings can be easily influenced by human error.

3.2 Safety considerations

In 2006, near London Waterloo station, a welder finished working at the switch shown in Figure 4 and undertook the inspection procedure of checking the switch for derailment hazards. The inspection failed to identify any derailment hazards. This was due to human errors that affected the measurement.

Two days later a loaded passenger train derailed while running over the switch.



source: http://www.raib.gov.uk/cms_resources.cfm?file=071218_R442007_Waterloo.pdf

Figure 5. Derailment at London Waterloo on 24th October 2006

The derailment on the 24th of October 2006 [2] is not unique; similar accidents involving derailed trains have been investigated. In the same year, a train derailed on the 11th of September near the same train station [2] and the following year, in 2007, a train derailed near Glasgow Exhibition Centre station [3]. The two main causes of these three train derailments are:

- 1) a grinding repair that failed to correct an existing rail profile defect or created another, and;
- 2) a incorrect inspection which failed to indicate the real condition of the part.

In January 2012, at a meeting with two Network Rail engineers, issues relating to the maintenance process of crossings were discussed. It was argued that, in Great Britain, there is a lack of understanding of how welds and grind repairs should be carried out. This is due to the fact that it is difficult to identify the exact shape and position to which the switch rails and the crossing must be repaired. The maintenance team which needs to repair the profile of switch rails and crossings are given little or no information at all on how the shape of switch rails and crossings must be altered. As a result, the condition of some crossings can become worse after the repairs have been carried out. Due to the higher forces that then take place in the defective repaired crossings, their life will be shortened, which has a negative impact on the life-cycle cost and the level of safety will decrease.

3.3 The requirements

It has been concluded that in Great Britain there is a need for a lightweight device that can:

- 1) advise the maintenance team on how the S&C should be repaired by means of welding and grinding;
- 2) carry out a specific profile inspections on switches and crossings, and;
- 3) do it efficiently.

Part of the first requirement is to advise the maintenance team which part, where and how much should it be rectified. The later requirement can be expressed more clearly by a set of five secondary requirements which are:

- 1) it should not be a expensive solution;
- 2) it should not need to schedule a train;
- 3) it should not need a track possession to be issued;
- 4) it should not need significant preparations in advance, and;
- 5) it should be available to be used at any given time.

3.4 The type of profile scan

In order to achieve this, it is necessary to acquire profile data from the parts that need to be inspected. Profile data can be acquired as separate 2D transversal slices or as a complete 3D model.

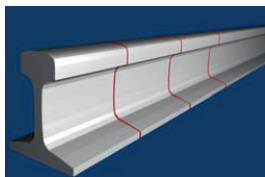


Figure 6. Consecutive 2D scans for the assessment of rails

The inspection of the switch rail and stock rail can be done by acquiring just 2D profile data. This is true because the inspection standards require the assessment of the rails in fixed locations along their length (e.g. assessed every 15 mm). Using the 2D profiles, the repair requirements can also be assessed in fixed locations along the rails. Due to the fact that the wear has a

similar pattern along the length of the rails, the repair requirements between two consecutive 2D profile scans are very similar. This allows for the repairs to be carried out between the 2D scans using just the data provided by the 2D profile scans.

The inspection and assessment of crossings cannot be done using just 2D transversal scans, as is the case for switches. This is because the Network Rail inspection standards specify the use of a one-metre (or two-metre) edge which is placed in three different orientations on the crossing in order to assess it for wear.

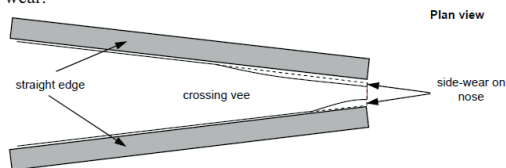


Figure 7. Representation of a crossing inspection task using two straight edges

In order to comply with the inspection requirements the scanning would have to be done either by positioning the line lasers in three different orientations, or by acquiring a 3D model of the crossing. The first solution is not attractive since lasers show a trade-off between the length of the scan and the precision of the acquired data (therefore a two-metre scan would impose low precision data) and the system would not be flexible for accommodating future inspection standards.

3D profiles can be achieved by combining 2D profiles of the part which needs to be inspected, but this comes with a loss in precision. This is because a moving object has six degrees of freedom (up/down, back/forward, left/right, roll, pitch and yaw). A scanning device which moves longitudinally along the rails, back and forth, will also be affected by the other five types of movement. When the 2D profiles are aligned, unless the movement can be constrained, it will introduce errors in the 3D model if they are not compensated.

An example of such error is the up and down movement of the recording device. This can be caused by wheel out of roundness, dirt between the wheel and rails or even imperfect rails. An undesired vertical movement (e.g. 0.5 mm up movement) can affect the measurements by up to several times the undesired movement (e.g. a resulting error of four times the undesired movement, 2 mm). This is usually the case when a laser beam makes a very small angle with the surface of the measured object (the laser beam is far from being normal on the surface of the object).

The proposed method does not present a solution for this matter and more research will be carried out to minimize or eliminate these errors.

It was found that the standards which set out the S&C inspection procedure in Germany involve similar practices to the ones in Great Britain. The switch profile is inspected by the use of a manual gauge which is placed on the part which needs to be inspected. Although the gauges differ, the principle is similar. The main railway infrastructure owner in Germany, Deutsche Bahn, have also identified the need to have better information regarding

the profile of crossings in order to carry out better repairs by means of welding and grinding.

4. CURRENT SOLUTIONS

4.1 Switch inspection wagon

Eurailsout is a company based in the Netherlands which is known in the railway industry. They produced the “Switch Inspection and Measurement” (SIM) [4] which is a wagon that can be pulled or pushed on the track to carry out various inspections on S&C. The wagon has an inspection system, a measurement system and an inertial measurement unit.

The inspection system is used to remotely visually inspect S&C and it does this by using video cameras mounted beneath it.



source: http://www.eurailsout.com/5-measuring-trains-and-vehicles_en.html

Figure 8. Switch Inspection and Measurement wagon

The measuring system is based on laser scanners which are able to measure the profile of the rail at each 20 mm while the wagon is moving at a speed of 40 km/h. It is believed that the SIM can measure several rail track and rail parameters including horizontal rail wear, vertical rail wear and track gauge. The measurements are done by using the profile data. Eurailsout are currently redeveloping their system to enhance its measuring capabilities; the development process is in the final stage, validation and acceptance. At the time of writing this paper, the abilities and limitations of the SIM to inspect and record 3D models of crossings were not clear.

The SIM proves to be a comprehensive tool for inspecting S&C, but when benchmarked against the five secondary requirements of this application, the following can be noted:

1. it may be costly (-);
2. it needs to be scheduled (-);
3. does not require track possession (+);
4. system availability may be an issue if two sets of S&C need to be inspected and they are a large distance apart (-);
5. There may be delays due to lack of availability (-).

Therefore, the authors believe that the SIM is not particularly well suited as a solution for this application.

4.2 Switch inspection vehicle

ZETA-TECH is a company based in New Jersey which has produced the “Automatic Switch Inspection Vehicle” (ASIV) [5] that is used in the USA to inspect S&C. As per its name, the ASIV is a vehicle that can drive on both road and rail and it uses laser scanners to measure the profile of the S&C.

The general principle of measurement is similar to the one adopted by Eurailsout. The car has a set of lasers which are able

to take 2D slices of the rails’ profile. The system is able to carry out certain measurements according to the British Network Rail standards. In order to do this, the ASIV has a set of software models which simulate manual gauges used in Great Britain. The gauges are virtually placed on the 2D rail slices and the result is computed.



source:

http://www.arena.org/files/library/2011_Conference_Proceedings/Development_and_Implementation_of_Automated_Switch_Inspection_Vehicle.pdf

Figure 9. Automatic Switch Inspection Vehicle

This vehicle had been demonstrated in Great Britain and it is able to correctly inspect switches against several Network Rail inspection requirements, as specified by the NR/L2/TRK/0053 standard.

As written in a report published on the AREMA (American Railway Engineering and Maintenance of Way Association) website “the ASIV rail profile data is then used to develop 3-D composite images of the turnout and its key components”, it is believed that the vehicle generates 3D images using 2D profile scans. At the time of writing this paper it was not clear to the authors whether or not compensation is made in order to counteract the errors which are introduced by the unrestricted movement of the vehicle.

The following can be noted when the ASIV is benchmarked against the five secondary requirements of this application:

1. it may be costly (-)
2. it does not need train scheduling (+)
3. it would probably require track possession (-)
4. the vehicle must be driven to the S&C (-)
5. it takes time to get it on the S&C (-)

Therefore, the authors believe that this technology is also not particularly suited for this application.

5. PROPOSED SOLUTION

5.1 Research hypothesis

This research tries to define an inspection system which could be successfully used to (1) aid welders on how to repair S&C, and (2) carry a NR/L2/TRK/0053 & NR/L2/TRK/1054 inspection while also not infringing on the five secondary requirements.

5.2 Description of proposed solution

The proposed solution is a lightweight trolley that can be manually pushed over the switch. It has a pair of wheels on each

side, the profile of a train wheel. The weight of the trolley is estimated at 15 Kg, although due to health and safety the maximum permitted weight is 25 Kg. A tachometer will be used so that the position of the trolley can be relatively monitored.

The main advantage of this approach is that the system will be used as a manual tool. This assures a good deal of flexibility and helps to meet the five secondary requirements.

At one end of the trolley, two line lasers will be used to scan the rail's profile. The criteria for the selection of the laser scanners takes into account the following: width of the scan, number of samples per one line scan, precision, sampling frequency, size, power consumption, robustness, price and overall quality. The chosen laser for this application is the scanCONTROL 2700 – 100 [6] produced by Micro Epsilon.



source: <http://www.micro-epsilon.co.uk/laser-scanner-profile-sensor/Laser-scanner-selection/index.html>

Figure 10. scanCONTROL line laser from Micro-Epsilon

This will allow the scanning of one switch half set (one stock rail and one switch rail which is on the same side) on each run. In Great Britain, most often just one half of the switch is required for inspection and repairs. The two rails (switch rail and stock rail) will be inspected at discrete points along their length. This will be done by computing 2D profile scans of the inspected parts. The crossing will be completely scanned by pushing the trolley two times over the crossing, one time on the straight path and one time on the diverging path. The inspection standards require that a 3D profile of the crossings must be scanned. Further research will be carried out to identify the best way of acquiring 3D profiles of crossings while reducing or eliminating the errors which account for the unconstrained movement of the trolley.

The power required by two laser scanners is just over 10 W, which will be provided by using rechargeable batteries.

The trolley will have a tablet computer and the software will be written in LabVIEW programming language. This decision is based on both the recommendation received from Micro Epsilon, as well on the advantages and flexibility that National Instruments offer. The operating system will be a variant of Windows for greater compatibility. The communication between the lasers and the tablet will be run through Ethernet, as it is easier to interface to a tablet computer.

The user will be able to inspect different parts of the switch and receive advice on how the profile should be rectified. The system will also inform the user if the profile is in accordance with the two Network Rail standards.

If assessed against the five secondary requirements the following can be noted:

1. less expensive (+)

2. no need to schedule a train (+)
3. no track possession is required (+)
4. no significant preparations required (just batteries need to be charged) (+)
5. no need to wait; the system can be used immediately (+)

The authors believe that the proposed solution offers a number of advantages over other systems which are currently commercially available.

6. CONCLUSION

The issue of profiling the S&C and supplying the maintenance team advice on how the repair should be carried out was considered and an alternative solution has been proposed.

It is believed by the authors that the proposed solution has many advantages when compared with other solutions which are currently available; this is mostly due to its small size which makes it easy to transport and use.

The proposed solution is an initial design that should accomplish what is desired, but it must be noted that it is not yet developed and tested and further research needs to be carried out.

7. ACKNOWLEDGMENTS

The authors thank the sponsors of AUTOMAIN for their financial support.

8. REFERENCES

- [1] International Union of Railways. (2011). "UIC Project Switches and Crossings; Inspection of Switches and Crossings State of the Art Report." Preliminary report.
- [2] Rail accident investigation branch, DfT. (2007). "Derailments at London Waterloo, 11 September and 24 October 2006." Available at: http://www.raib.gov.uk/cms_resources.cfm?file=/071218_R44_2007_Waterloo.pdf, accessed: 15 August 2012.
- [3] Rail accident investigation branch, DfT. (2009). "Derailment near Exhibition Centre station, Glasgow, 3 September 2007." Available at: http://www.raib.gov.uk/cms_resources.cfm?file=/090212_R04_2009_Glasgow.pdf, accessed: 15 August 2012.
- [4] Eurailsout. "Swich Inspection and Measurement fact sheet." Available at: http://eurailsout.com/index.php?option=com_docman&task=doc_download&gid=13&Itemid=15&lang=en, accessed: 15 August 2012.
- [5] ZETA-TECH. (2011). "Development and Implementation of Automated Switch Inspection Vehicle." Available at: http://www.arena.org/files/library/2011_Conference_Proceedings/Development_and_Implementation_of_Automated_Switch_Inspection_Vehicle.pdf, accessed: 15 August 2012.
- [6] Micro-Epsilon. "Laser Profile Scanner." Available at: <http://www.micro-epsilon.co.uk/laser-scanner-profile-sensor/Laser-scanner-selection/index.html>, accessed: 15 August 2012.

Data to Decision through Contextual Presentation of Railway Infrastructure Performance

Christer Stenström

Division of Operation, Maintenance
and Acoustics
Luleå University of Technology
Luleå, SE-971 87, Sweden
+46 920 49 14 76

christer.stenstrom@ltu.se

Karina Wandt

Division of Operation, Maintenance
and Acoustics
Luleå University of Technology
Luleå, SE-971 87, Sweden
+46 920 49 24 83

karina.wandt@ltu.se

Aditya Parida

Division of Operation, Maintenance
and Acoustics
Luleå University of Technology
Luleå, SE-971 87, Sweden
+46 920 49 14 37

aditya.parida@ltu.se

ABSTRACT

Asset performance measurement is an integral part of business process designed to support infrastructure managers in decision-making. Managing infrastructure performance throughout an asset's life cycle is a challenge; assessment is a complex issue involving various inputs and outputs, as well as conflicting requirements of different stakeholders. For railway infrastructure managers, the most critical issue is to reduce maintenance possession time, to minimise train delays and meet passengers' and societal needs. In this study, we collect and analyse work order and train delay data for one section of the iron ore railway line in Sweden. The aim is to present the overall performance of an asset for its end users, considering both the asset context and the user context, to ensure effective planning and decision-making. The case study results can also be used for internal and external benchmarking, and to identify performance problems of the infrastructure.

Keywords: Performance measurement, composite indicators, data quality, maintenance, eMaintenance, railway, train delay, failures, heavy haul, context

1. INTRODUCTION

Physical assets are a valuable part of the overall business process for which companies treats asset performance measurement and management an integral part of their strategic planning and policy making. In asset-intensive industries, the maintenance costs of assets constitute a significant amount of the total business cost. Non-availability and/or downtime of an asset will have an impact on the plant and asset capacity, product quality, and cost of production, as well as on health, safety and the environment. Asset performance measurement and management is a multi-disciplinary process; it provides critical support to heavy and capital-intensive industries by keeping assets, e.g. machinery and equipment, in a safe operating condition. It is generally accepted that asset performance management is key to the long-term profitability and sustainability of an organization.

Due to today's competitive global business environment, asset utilisation and performance optimisation throughout an asset's life cycle are important to asset owners and infrastructure managers (IMs). Physical assets represent the basic infrastructure of all businesses and their effective management

is essential to meet business objectives and goals. Thus, it is necessary to plan, monitor and control assets' performance throughout their entire life cycle, from the development/procurement stage to their eventual disposal. Life cycle costing seeks to optimise value for asset owners and infrastructure managers; to this end, it considers all the cost factors of the asset during its entire operational life. Today, with available knowledge and technology, it is perceived that "asset performance measurement and management can be planned and controlled," thus, it is essential for management to understand and calculate an asset's availability and capacity utilisation to make effective repair and replacement decisions.

Health monitoring of the infrastructure or of any individual asset is a critical issue; management must provide right information on an asset's health status to achieve better performance and reliability, safely and with minimum costs. The advancement in information technology has had a significant impact on the asset management information system; we are now better able to determine an asset's health status, thereby supporting good decision-making. Technological advancements, including embedded and wireless sensors, automated controls and data analysis management, have led to new and innovative methods in asset health monitoring. Further, rapid growth in networking systems, especially through the internet, has overcome the barriers of distance, allowing real time data transfer to occur easily from different locations [27]. With the emergence of intelligent sensors to measure and monitor the health state of a plant and its machines and with the gradual implementation of information and communication technologies (ICT) in organizations, the conceptualization and implementation of eMaintenance is becoming a reality [21]. Improved connectivity, faster transfer of data and the ability to store and analyse large amounts of data are now required by maintenance managers. Current eMaintenance tools have already utilised existing web and computing network technology to form a maintenance infrastructure for integrating and synchronising various maintenance information, supporting and enhancing collaboration between different users [16].

Corporate strategy dictates how to achieve business objectives and create value for the stakeholders. However, without a comprehensive description of strategy, executives cannot easily communicate the strategy among themselves or to their employees [8]. The management of an organisation must

convert the corporate strategy and objectives into specific objectives for each of the organisation's various hierarchical levels. An appropriate asset performance measurement system in an organisation has become a necessity, as without such assessment, it is not possible to attain the desired objectives.

A number of asset performance measurement frameworks have been developed applying the "Balanced Scorecard" [10] approach to ensure that all operational and maintenance activities of assets are aligned with an organisation's corporate strategies and objectives in a balanced manner. For details, see Parida and Chattopadhyay [20]. Parida and Chattopadhyay have developed a multi-criteria hierarchical maintenance performance measurement framework, which meets the organisational requirements of both external and internal stakeholders and identifies the various performance indicators (PIs) from a balanced and integrated view point. The framework has been tested in three Scandinavian industries with modifications to suit their individual organisational requirements.

In this study, after discussing the importance of performance measurement, work orders (WOs) and train delay data are collected and analysed for the iron ore railway line in Sweden. The aim is to present the overall performance of assets for the end users, considering both the asset context and the user context, to ensure effective planning and decision-making (Figure 1). The case study results can also be used for internal and external benchmarking, and to identify performance killers [25] of the infrastructure. Asset context and user context can be described as follows:

- The physical asset performance context: the asset's overall performance is determined using necessary performance indicators (PIs). In advanced setups, learning of PIs pattern for failure can be used in future prediction of similar events.
- The user context: asset performance with the right PIs, in the right format, is delivered to the right person, at the right time.

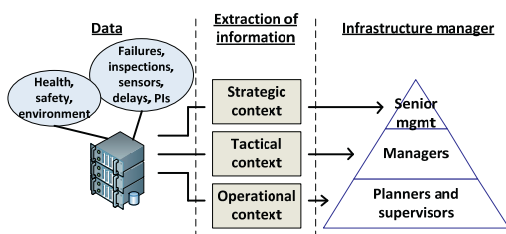


Figure 1. Visualisation of the need of a context based presentation of asset performance for business stakeholders

The structure of this paper is as follows. After introducing the topic in section 1, we discuss infrastructure performance measurement for decision-making in section 2. Section 3 deals with data quality issues in the decision process. The iron ore railway line case study and results are presented in section 4, followed by the discussion and conclusions.

2. INFRASTRUCTURE PERFORMANCE MEASUREMENT

- *What is infrastructure and why is PM required for infrastructure?*

Due to the increasing awareness that maintenance not only ensures safety and track performance, but creates additional value in the business process, Trafikverket (Swedish transport administration) is treating maintenance as an integral part of its business process, i.e. applying a holistic view to the infrastructure maintenance process to meet customer requirements [12]. Its infrastructure maintenance process visualises both front and back end processes in track maintenance [2]. One front end process is determining track maintenance demands, supported by measures such as track capacity and track quality [2].

From the infrastructure management perspective, achieving a punctual and cost-effective railroad transportation system requires ongoing development in maintenance engineering. Cost-effective maintenance processes help to achieve budget targets, while punctuality is required by stakeholders [1].

Companies are now using scorecards to manage their strategy over the long term in a number of critical processes [9] including the following:

1. Clarify and translate vision into objectives and strategy;
2. Communicate and link strategic objectives with the performance measures at different hierarchical levels;
3. Plan and set targets linked with KPIs/MPIs and aligned with strategic initiatives;
4. Enhance strategic and performance feedback and learning.

The KPIs translate aggregate measures from the shop floor to the strategic level. The real challenge lies in measuring all the KPIs; some are difficult to measure since they are intangible and qualitative in nature. Organizations need a framework to align their performance measurement system with the corporate strategic goals of a company by setting objectives and defining key performance at each level [14]. The performance measurement which forms part of the asset performance measurement system needs to be aligned with the organizational strategy [17]. The PIs must be considered from the perspective of the multi-hierarchical levels of the organization. As per [17], maintenance management must be carried out in both strategic and operational contexts and the organizational structure generally comprises three levels. Three hierarchical levels in most firms are the strategic or top management level, the tactical or middle management level, and the functional/operational level [20]. Two major strategic requirements of a successful corporate strategy relevant for the performance assessment are (Figure 2) the following:

1. Cascading the objectives down from the strategic level to the shop floor;
2. Aggregating performance measurements from the shop floor up to the strategic level.

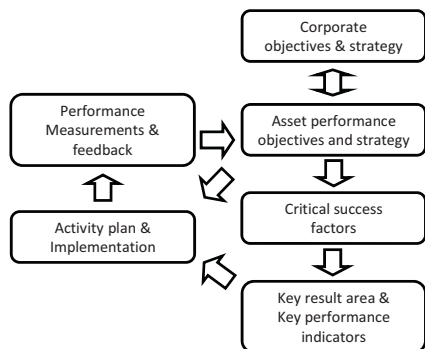


Figure 2. Strategic maintenance performance measurement process [19]

Results are visualised in key result areas (KRAs); critical success factors (CSFs) are those required to achieve the objectives of the KRAs.

2.1 Cascading down to the shop floor from the strategic level

The strategic objectives are formulated based on the requirements of the stakeholders, both internal and external. The plant capacity and resources are considered from a long-term point of view and are matched with each other. These strategic or corporate objectives cascade down the hierarchical level of the organization through the tactical level which considers the tactical issues, such as financial and non-financial aspects, both from the effectiveness and the efficiency point of view. The bottom level includes the shop floor engineers and operators.

2.2 Aggregating performance assessments from the shop floor up to the strategic level

The performance at the shop floor level is measured and aggregated to evaluate whether the corporate objectives have been achieved. Inspections, physical measurements or sensor/condition based measurements generate data which are analysed through programming or simulation, facilitating effective decision-making at the managerial/strategic level. The adoption of appropriate processes is vital to successfully align performance measurement to objectives. The energy and creativity of committed managers and employees needs to be harnessed to drive the desired organisational transformations [28]. This, in turn, leads to the empowerment of employees in the organization.

3. DATA QUALITY

For recording and conveying information about the maintenance management process, a maintenance documentation system is an essential operational requirement. Maintenance documentation can be described as any record, catalogue, manual, drawing or computer file containing information that might be required to facilitate maintenance work [13]. For

maintenance, failure records are especially relevant; thus, failure data need to be recorded in a way that allows further computational analysis. A uniform definition of failure and a method of classifying failures are essential when data from different sources (plants and operators) are combined in a common maintenance database.

The process that begins by collecting the data and ends by presenting the information to the end user can be described as an interactive decision-making process. It can be divided into the following five steps: data collection, data transition, data fusion, data analysis and data presentation [29]. Each phase is important, as each affects the quality of the data.

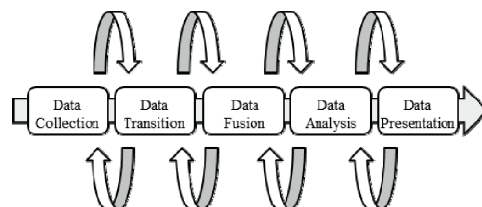


Figure 3. Generic decision making process, adapted from [29]

The lack of relevant data and information is one of the main problems for decision-making within the maintenance process [21]. The provision of the right information to the right user with the right quality and at the right time is essential [11, 21]. Data must be essential and relevant for its specific user; high-quality data are commonly defined as data that are appropriate for use by data consumers [26, 30]. Hence, to provide high quality data to the data consumer, one must understand what quality means to those who use the data [30].

Wang [30] presents a framework of data quality consisting of four categories: intrinsic, contextual, representational and accessibility. Each category relates to different data quality dimensions, as described in Table 1.

Table 1. Data quality aspects [30]

Category	Dimension
Intrinsic	Believability, Accuracy, Objectivity, Reputation
Contextual	Value-added, Relevancy, Timeliness, Completeness, Appropriate amount of data
Representational	Interpretability, Ease of understanding, Representational consistency, Concise representation
Accessibility	Accessibility, Access security

The context of the task for which the data will be used is essential to determine what data quality means in a particular instance. A study of data quality relating different contextual quality issues to the dimensions described in Table 1 has observed the following underlying causes for data not supporting the intended tasks: (i) missing or incomplete data;

(ii) inaccurately defined or measured data; and (iii) data that could not be properly collected [26]. Figure 4 illustrates an example of how these contextual data quality issues can be linked to the different phases in the generic decision-making process, described above in Figure 3.

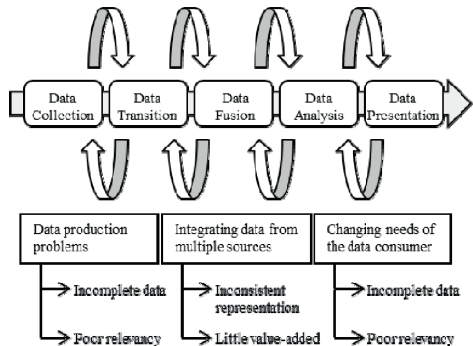


Figure 4. Data quality issues in a generic decision-making process, adapted from [26, 29]

4. CASE STUDY

The case study presents a railway asset’s overall performance, considering the context of the asset and the context of the end user. Data have been collected and analysed for indicators considered important for the asset’s performance. The results can facilitate efficient and effective decision-making for the operation and maintenance managers of the asset and can also be applied to the benchmarking of similar assets. The subject of the case study is a heavy haul railway line in Sweden.

The infrastructure and the IM are divided into three levels: strategic, tactical and operational, i.e. senior managers, middle managers and supervisors, respectively. Each level monitors the asset to meet its own needs (Figure 5). However, given the vast amount of data at the strategic level, the case study focuses on the tactical and operational levels. The asset morphology follows:

Corridor or network → Lines or routes → Sections → Zones → Systems, e.g. signalling system

The iron ore line is one line out of 84 lines in the Swedish network. Section 111 is one section out of 6 within the iron ore line, and there are 16 zones within that section. The management hierarchy of the asset takes the following form:

- Senior managers: Lines
- Middle managers: Sections
- Operational managers or supervisors: Zones and systems

Given three scorecards, strategic, tactical and operational, the case study focuses largely on the operational scorecard, paying some attention to the tactical scorecard.

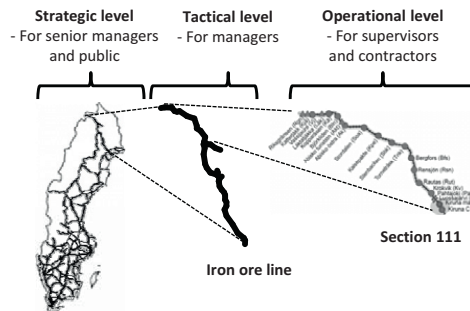


Figure 5. Railway network of Sweden, showing the subsystems and their stakeholders, each with specific needs for monitoring performance

4.1 Data collection and analysis

Railway section 111 of the iron ore line is a 128 km 30 tonne axle load mixed traffic heavy haul line stretching from Kiruna to Riksgränsen, on the Norwegian boarder (Figure 6). The data presented in Table 2 are considered important for a context focused presentation of railway infrastructure performance.

Table 2. Data for analysis

Data	Background
Overview of the asset	Important to know the operational environment and how it affects maintenance planning etc.
Sections and zones of the line	The sections of the line can have different subsystems, e.g. number of switches and crossings, and therefore the performance can vary
Length of each section/zone, and number of switches and crossings (S&C)	Number of failures in a section or zone can depend on the length of the section/zone or on the number of switches and crossings
Corrective maintenance work orders (WOs), i.e. failures	Failure frequencies of systems, etc.
Train delay (due to WOs)	Severity
Maintenance time	Important for administrative, logistic and repair time analysis and planning

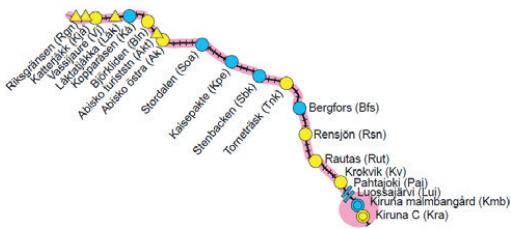


Figure 6. Zones of section 111, stretching from Kiruna city to Riksgränsen, the border of Norway

Corrective maintenance WOs were collected for 2001.01.01 - 2009.12.01, i.e. 8 years and 11 months. Out of 7 476 WOs in total, 1 966 mention train delays, i.e. 26 %. However, the train delay data have a skewed distribution with some long delays resulting in a long tail. The two percent with the longest delays are therefore considered as outlier cases. Outliers are preferably analysed before decision-making, but this is beyond the scope of this research. In fact, some railway infrastructure assets tend to give very long delays when it fails, e.g. the overhead contact system due to tear down of the contact wire by locomotive pantographs; alternatively the outliers could have been discounted per asset type. However, the analysis is based on WOs with delays up to the 98th percentile. In terms of WOs, 1 926 out of 1 966 WOs are considered; in terms of delay, this represents 112 616 minutes out of 166 693 minutes.

The corrective maintenance work order data consist of urgent inspection remarks reported by the maintenance contractor, as well as failure events and failure symptoms identified outside the inspections, commonly reported by the train driver, but occasionally reported by the public. The work order failure reports include the three categories of RAM (reliability, availability and maintainability) failure as identified by the European Standards EN 50126 [4], see Figure 7. Immediate action is required, if the fault negatively influences safety, causes train delay, or affects a third party or the environment.

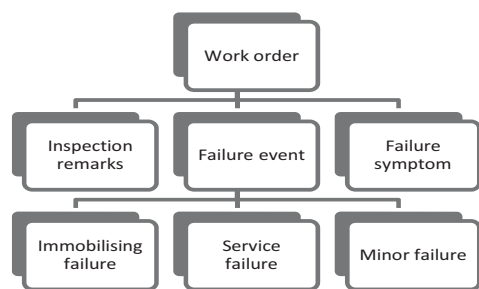


Figure 7. Work order description. The three RAM (reliability, availability, maintainability) categories are: immobilising failure, service failure and minor failure [4]

Indicators extracted from the data are presented in Table 3 below.

Table 3. Indicators calculated

Indicator	Comment
Number of WOs [No.]	Corrective maintenance work orders, i.e. functional failures
Train delay [min]	Delay caused by the WOs
Maintenance time	Consists of administrative time, logistic time and active repair time. For details, see [3, 6]
Risk [a.u.]	Risk is a composite indicator [18, 22] of WOs and train delay, defined as:
	$Risk = (a\alpha^2 + b\beta^2 + c\gamma^2)^{1/2}$, $a, b, c \in \mathbb{R}$
	where α = WOs, β = train delay, γ = maintenance time. The indicator can be seen as a measure of business risk. For further background, see probability-consequence diagram [7] or failure mode effect analysis.

4.2 Results

From the collected and analysed data, we find performance results for the operational and tactical levels. Table 4 and Figure 8 give the performance for the various zones of the railway section (Figure 6). The zones in the table correspond to the points in Figure 6, including the track metres to the next zone, e.g. zone Kv = Point Kv + track to Rut. Table 5 gives the top three systems with the highest risks, i.e. the performance killers. Table 6 shows the tactical level of section 111. However, the iron ore line consists of six sections; one is considered.

A similar study of the iron ore line carried out by [5] is useful for comparative purposes.

Table 4. Railway zones performance for the operational level. Some noteworthy values are marked in grey and bold

Zone	Length [m]	S&C [No.]	WOs /length [No./km]	Delay /length [min/km]	Maint. time (median) [Min]
Kmb (Pea)	4394	1	20	1396	157
Kv	9039	7	28	1646	135
Rut	8869	3	12	680	160
Rsn	7778	3	12	583	144
Bfs	8907	3	21	1147	151
Tnk	8477	4	23	777	129
Sbk	7929	6	15	700	198
Kpe	10612	3	16	716	148
Soa	8974	3	11	436	108
Ak	912	12	122	3063	112
Akt	6562	0	5	260	151
Bln	7164	3	25	1584	136
Kå	6742	4	13	1361	182
Låk	1130	0	19	1617	202
Vj	4088	4	36	2662	227
Kjå-Rgn	1767	0	8	1285	195
Whole sec.	103344	56	18	1020	153

Table 5. Railway zone systems performance for the operational level. Some noteworthy values are marked in grey and bold. Risk = $(\alpha^2 + (100^{-1}\beta)^2 + \gamma^2)^{1/2}$

Zone	Top three systems in terms of risk (no.)	WOs [No.]	Delay [min]	Risk rank [a.u.]	Risk rank /length [a.u./km]
Kmb (Pea)	Track	58	4731	75	17
	Pos. Sys.	11	175	11	2,5
	Fault disappeared	10	150	10	2,3
Kv	S&C (7)	74	4443	86	-
	Fault disappeared	64	1072	65	7,2
	Track	32	3435	47	5,2
Rut	Fault disappeared	19	231	19	2,2
	Track	14	1170	18	2,1
	Pos. Sys.	15	421	16	1,8
Rsn	Fault disappeared	18	236	18	2,3
	Track	17	588	18	2,3
	Pos. Sys.	16	564	17	2,2
Bfs	Fault disappeared	58	781	59	6,6
	Track	42	3439	54	6,1
	Pos. Sys.	31	1353	34	3,8
Tnk	S&C (4)	55	1648	57	-
	Fault disappeared	47	560	47	5,6
	Pos. Sys.	32	862	33	3,9
Sbk	Fault disappeared	21	191	21	2,7
	Signalling	20	481	21	2,6
	S&C (6)	20	348	20	-
Kpe	S&C (3)	46	1743	49	-
	Signalling	35	820	36	3,4
	Track	23	1257	26	2,5
Soa	Fault disappeared	30	408	30	3,4
	S&C (3)	20	734	21	-
	Pos. Sys.	12	459	13	1,4
Ak	S&C (12)	37	833	38	-
	Fault disappeared	26	363	26	29
	Track	13	255	13	15
Akt	Pos. Sys.	7	160	7	1,1
	Track	4	194	4	0,7
	Fault disappeared	4	110	4	0,6
Bln	S&C (3)	40	1013	41	-
	Fault disappeared	40	821	41	5,7
	Signalling	13	2876	32	4,4
Kâ	Pos. Sys.	15	1546	22	3,2
	Track	11	1558	19	2,8
	S&C (4)	16	368	16	-
Lâk	OCS	5	734	9	7,9
	Pos. Sys.	5	479	7	6,1
	Track	3	401	5	4,4
Vj	S&C (4)	55	2966	62	-
	Signalling ctrl	16	1207	20	4,9
	Pos. Sys.	17	720	18	4,5
Kjâ-Rgn	Signalling ctrl	4	834	9	5,2
	Sectioning station	1	513	5	3,0
Rgn	Track	2	442	5	2,7

a.u. = arbitrary unit

Table 6. Railway section performance for the tactical level. Risk = $(\alpha^2 + (100^{-1}\beta)^2 + \gamma^2)^{1/2}$. One section out of six is presented.

Section	Length [m]	S&C [No.]	WOs /length [No./km]	Delay /length [min/km]	Risk rank /length [a.u./km]	Maint. time (median) [Min]
111	103344	56		18	1020	21
Top three systems in terms of risk						
			WOs [No.]	Delay [min]	Risk rank [a.u.]	Risk rank /length [a.u./km]
			404	16880	438	-
			308	28590	420	3,28
			396	5876	400	3,13

a.u. = arbitrary unit

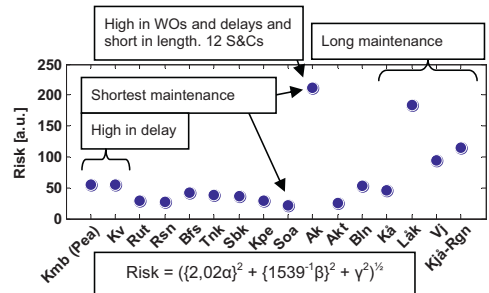


Figure 8. Risk ranks of the zones, taking into account WOs (α), train delay (β) and maintenance time (γ). Constants chosen to normalise parameters to give the same weight on the risk, i.e. equal weighting. Risks are divided by the length of the zones.

5. DISCUSSION

Various data and indicators are identified as important for presenting railway infrastructure performance in its context. However, railway infrastructure has a wider context than the one presented in this study, e.g. preventive maintenance and cold climate effects [24]. Moreover, the analysis comes from data over a nine year period, but aggregating data over nine years does not necessarily provide accurate information of the present state. Thus, analysis of shorter timespans is also important and needs to be considered [23].

Risk ranks are calculated in different ways for the section, zones and systems (Tables 5 and 6, Figure 8), by use of different weights and considering two or three parameters. It shows its variability but also constrains of the various levels, e.g. track system failures (WOs) may fit to be divided by zone length, while S&C does not. Moreover, the weights in Figure 8 were calculated from the 16 zones, such calculation was not possible in Table 6 since only one section was considered. It should also be noticed that failures, train delay and maintenance time are functions of each other to some extent.

The case study shows how a physical asset's performance can be presented in its unique working context. However, even if descriptive statistics give valuable information for decision-making, it is lagging indicators. Simple and multiple linear regressions have been carried out to predict the train delays and number of work orders, but it provided weak results, requiring further work in this area.

The quality of the data has not been the main focus in the presented case study. However, in order to gain high quality data, one must understand what quality means to those who use the data [30]. The context of the task, i.e. where the data are used, is essential to determine data quality and this case study can be used as a starting point for further research in data quality issues related to contexts.

Contextual data quality emphasizes that the data must be relevant, timely, complete, and appropriate in terms of amount in order to add value. However, achieving contextual data quality is a subject for future research, since contexts and tasks vary over time and between data consumers [15]. Letting the data consumer parameterize the contextual dimensions (presented in Table 1) for each task is a possible approach [30].

6. CONCLUSIONS

This study takes some initial steps in presenting railway infrastructure performance while considering both the asset context and the user context, to facilitate efficient and effective decision-making. A composite indicator has been constructed to summarise the overall performance of a complex asset into a single number, easier to interpret by decision-makers than presenting multiple indicators and plots (Figure 8). However, parameter correlation, expert opinion weighting and sensitivity analysis are future work to consider. Moreover, further work is also needed for prediction and for taking into account more indicators to create a broader context.

7. ACKNOWLEDGMENTS

The authors would like to thank Adjunct Prof. and Dr. Per-Olof Larsson-Kräik and Dr. Ulla Juntti, Luleå Univ. of Tech. (LTU), for their assistance and expertise in railways. The authors would also like to thank Dr. Diego Galar and Stephen Famurewa, LTU, for valuable discussions. The research has been supported and funded by Luleå Railway Research Centre (JVTC), Trafikverket (Swedish Transport Admin.), and the European project Bothnian Green Logistic Corridor (BGLC); the authors would like to thank them for their support.

8. REFERENCES

- [1] Åhrén, T., "Maintenance Performance Indicators (MPIs) for Railway Infrastructure: Identification and Analysis for Improvement", Division of Operation and Maintenance Engineering, Luleå University of Technology, 2008.
- [2] Banverket, "Annual report 2004", Banverket, Borlänge, 2005.
- [3] CEN, "EN 13306: Maintenance Terminology", European Committee for Standardization, Brussels, 2011.
- [4] CEN, "EN 50126: Railway Specifications – The Specification and Demonstration of Reliability, Availability, Maintainability and Safety (RAMS)", European Committee for Standardization, Brussels, 1999.
- [5] Famurewa, S.M., Asplund, M., Parida, A., Kumar, U., "Railway maintenance performance measurement: Perspective for improvement", D. Baglee, ed., In: MPMM 2012: The 2nd International Congress on Maintenance Performance Measurement & Management Conference Proceedings, 12-13th Sep 2012, University of Sunderland 2012.
- [6] IEC, "IEC 60050-191: International Electrotechnical Vocabulary: Chapter 191: Dependability and quality of service", International Electrotechnical Commission (IEC), Geneva, 1990.
- [7] ISO/IEC, "ISO/IEC 31010: Risk Management – Risk Assessment Techniques", International Organization for Standardisation (ISO), Geneva, 2009.
- [8] Kaplan, R.S., Norton, D.P., "Strategy Maps: Converting Intangible Assets into Tangible Outcomes", 1st edn. Harvard Business School Press, Boston, 2004.
- [9] Kaplan, R.S., Norton, D.P., "The Balanced Scorecard: Translating Strategy into Action", 1st edn. Harvard Business Press, Boston, 1996.
- [10] Kaplan, R.S., Norton, D.P., "The Balanced Scorecard - Measures that Drive Performance", Harvard Business Review, 70(1), pp. 71-79, 1992.
- [11] Karim, R., Candell, O., Söderholm, P., "E-maintenance and information logistics: Aspects of content format", Journal of Quality in Maintenance Engineering, 15(3), pp. 308-324, 2009.
- [12] Karlsson, V., "An overall view of maintenance", European Railway Review, 11(3), pp. 11-17, 2005.
- [13] Kelly, A., "Maintenance Systems and Documentation", 1 edn. Butterworth-Heinemann Ltd, UK, 2006.
- [14] Kutucuoglu, K.Y., Hamali, J., Irani, Z., Sharp, J.M., "A framework for managing maintenance using performance measurement systems", International Journal of Operations & Production Management, 21(1/2), pp. 173-195, 2001.
- [15] Madnick, S.E., "Integrating information from global systems: Dealing with the "on-and off-ramps" of the information superhighways", Journal of Organizational Computing, 5(2), pp. 69-82, 1995.
- [16] Muller, A., Crespo-Marquez, A., Iung, B., "On the concept of e-maintenance: review and current research", Reliability Engineering and System Safety, 93(8), pp. 1165-1187, 2008.
- [17] Murthy, D.N.P., Atrens, A., Eccleston, J.A., "Strategic maintenance management", Journal of Quality in Maintenance Engineering, 8(4), pp. 287-305, 2002.
- [18] OECD, EU, JRC - EC, "Handbook on Constructing Composite Indicators: Methodology and User Guide", Organisation for Economic Co-operation and Development (OECD), 2008.
- [19] Parida, A., Åhrén, T., Kumar, U., "Integrating maintenance performance with corporate balanced scorecard", O.P. Shrivastav, B. Al-Najjar and R.B.K.N. Rao, eds., In:

- Condition Monitoring and Diagnostic Engineering Management : COMADEM 2003 : proceedings of the 16th International Congress, 27-29 August 2003 2003, pp. 53-59, , Växjö universitet2003.
- [20] Parida, A., Chattopadhyay, G., "Development of a multi-criteria hierarchical framework for maintenance performance measurement (MPM)", *Journal of Quality in Maintenance Engineering*, 13(3), pp. 241-258, 2007.
- [21] Parida, A., "Maintenance performance measurement system: Application of ICT and e-Maintenance concepts.", *International Journal of COMADEM*, 9(4), pp. 30-34, 2006.
- [22] Saisana, M., Tarantola, S., "State-of-the-art report on current methodologies and practices for composite indicator development", EUR 20408 EN. European Commission-JRC, Italy, 2002.
- [23] Stenström, C., "Link and effect model for performance improvement of railway infrastructure, Licentiate thesis", Luleå University of Technology, 2012.
- [24] Stenström, C., Famurewa, S.M., Parida, A., Galar, D., "Impact of cold climate on failures in railway infrastructure", D. Baglee, ed. , In: *MPMM 2012: The 2nd International Congress on Maintenance Performance Measurement & Management Conference Proceedings*, 12-13th Sep 2012, University of Sunderland2012.
- [25] Stenström, C., Parida, A., Kumar, U., Galar, D., "Maintenance Value Drivers, Killers and their Indicators", D. Galar, A. Parida, H. Schunnesson and U. Kumar, eds. , In: *1st International Conference on Maintenance Performance Measurement and Management (MPMM)*, 13-15 Dec. 2011, pp. 125-130, , Luleå University of Technology2011.
- [26] Strong, D.M., Lee, Y.W., Wang, R.Y., "Data quality in context", *Communications of the ACM*, 40(5), pp. 103-110, 1997.
- [27] Toran, F., Ramirez, D., Casans, S., Navarro, A.E., Pelegri, J., "Distributed virtual instrument for water quality monitoring across the Internet", *Instrumentation and Measurement Technology Conference, 2000. IMTC 2000. Proceedings of the 17th IEEE 2000*, pp. 652-656 vol.2, 2000.
- [28] Tsang, A.H.C., "A Strategic Approach to Managing Maintenance Performance", *Journal of Quality in Maintenance Engineering*, 4(2), pp. 87-94, 1998.
- [29] Wandt, K., Karim, R., Galar, D., "Context adapted prognostics and diagnostics", *The Ninth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, 12-14 June 20122012.
- [30] Wang, R.Y., "Beyond accuracy: What data quality means to data consumers", *Journal of Management Information Systems*, 12(4), pp. 5-34, 1996.

Prognostic and Health Management of Wheel Condition: Integration of Wheel Defect Detection and Wheel Profile Monitoring Data

Matthias Asplund
 Operation, Maintenance and
 Acoustics, Luleå University of
 Technology
 97187, Luleå, Sweden
 +46 920 49 10 62
 matthias.asplund@ltu.se

Stephen M. Famurewa
 Operation, Maintenance and
 Acoustics, Luleå University of
 Technology
 97187, Luleå, Sweden
 +46 920 49 23 75
 stephen.famurewa@ltu.se

Matti Rantatalo
 Operation, Maintenance and
 Acoustics, Luleå University of
 Technology
 97187, Luleå, Sweden
 +46 920 49 21 24
 matti.rantatalo@ltu.se

ABSTRACT

The need for proactive maintenance is increasing within railway infrastructure management due to the capacity requirement. The business environment of railway transport could be described as a multi-stakeholder model which requires intelligent infrastructure with monitoring capability of the operation to facilitate sustainable transport system. The demand to support and extend the design life time of railway infrastructure as well as their dependability during the life is also another concern which requires adequate management of the maintenance process. An approach to meet this challenge and demands is implementing proactive maintenance strategy, in which prognostic health management of wheel condition is an integral part. This article presents different monitoring system deployed in the railway transport, the monitoring system include wayside and on-board equipment. A case study of the wheel defects records from the detection system on the iron ore line of the Swedish transport administration is presented where the trends are shown and analysed. The development of the wheel condition monitoring, from only wheel defects detection to both wheel defects detection and wheel profile measurement is discussed. Finally the need for e-maintenance solution, to facilitate the anticipated prognostic aspect in wheel condition measurement is elaborated. This will support effective maintenance decisions that are required for a competitive and sustainable transport system which in turn enhances inherent network capacity.

Keywords

e-maintenance, out-of-round, wheel flat, profile, prognostic health management.

1. INTRODUCTION

The maintenance trend in the railway industries emerged from basic policy of “fail & fix” through “predict & prevent” and then to the condition monitoring policy. The current trend is still emerging as more and more challenging concerns such data integration, mining, analysis and decision support are rising. This trend is driven by increasing demand for cost effectiveness, capacity, safety, and also multi-stakeholder structure of the railway industries. It is a lot of reorganisation and changes in railway maintenance, even though the fundamental objective of achieving excellence remains unchanged. The main concern of all

physical asset management including railway infrastructure is balancing performance, risk, costs and other inputs as shown in Figure 1 [6]. In relation to this fact, wheel condition monitoring system in a prognostic health management plan will obviously add to cost in a short term. On the other hand, the system will decrease the operational risk, enhance performance and in the long time contribute to cost reduction.

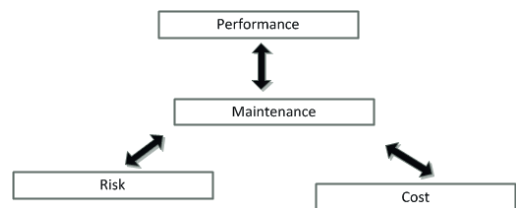


Figure 1. Balance mix in Maintenance excellence

Due to the inherent characteristics of the capacity of a railway system, the failure driven capacity consuming events within a railway network should be kept to a minimum. This could be achieved by the use of appropriate existing and new condition monitoring systems which can detect and predict failure events at an early stage.

A review of structural health monitoring in the railway industries has presented critical parameters relating to the condition of rolling stock and also relevance of monitoring system to both track owners and train operators [3]. Also a review on imperfections on wheel thread called out-of-rounds has described the root cause of the development of out-of-round (OOR) and the possible damage on either track or vehicle [12]. In the large spectrum of wheel imperfections wheel flat is known for severe repeated high-frequency impact loads. For instance, OOR wheels generate high force peaks that can damage the track. The high force from a wheel defect is detrimental for the track and the vehicle. The force degrades rail welds, ballast sleeper, fasteners and rail pads and other track components [2]. There are also some correlations in between wheel condition and track force. In addition since the force signature has been established to be different in between wheels in the same wagon [14], same might

perhaps apply to force signature of wheels with similar imperfections but different position in the wagon.

This is a big problem especially in locations with extreme climate conditions as the temperature makes the rail more brittle and more sensitive to impact that can damage the track.

Wheel failures appear most frequently during autumn and winter. Many of the wheel failures cause disturbance on the traffic, for example around 1000 h delay is attributed to wheel failures on the Swedish network. Furthermore, the socioeconomics costs, due to wheel failures on track are very high.

Basically, lot of focus on the health monitoring of railway systems are on wheels and bogies of rolling stock, since they amount for main proportion of maintenance cost for the operator. Nonetheless, way-side monitoring units are more preferable due to cost effectiveness, as it is expensive to mount monitoring devices on all vehicles [10].

Health monitoring systems in railway infrastructure can either be reactive or proactive in principle. But the focus of condition based maintenance requires a proactive monitoring system which gives room to adequate prognostics and preventive maintenance approach. Some examples of the reactive systems which are installed to limit damage are; dragging equipment detector, hot box detector and hot/cold wheel detector and sliding wheel detector. Although this category of equipment is useful but they have the limitation of not able to capturing potential failure in sufficient time before failure to allow for proactive decision making. Then some examples of the proactive systems are; vehicle inspections, hunting vehicle and bogie performance monitoring, wheel condition monitoring and acoustic bearing detectors. Those characterized by better possibility of understanding trends in deterioration of vehicle components and analysing the condition of the affected part or system for maintenance decision support [4, 10].

The wheel defect detector system is a reactive system and the wheel profile measurement is a proactive system which supports failure prognosis and effective maintenance decision. In many case when a flat wheel is detected the damage has already taken place, this require time consuming inspection of the track and sometimes the maintenance actions is needed. These events limit the traffic since the speed is reduced until the inspection is carried out. In most cases this inspection is manual, and the length of the track section to be inspected is decided from the traffic control centre [7]. This can be in between the detector that finds this defect and the preceding detector. Still one is not sure that no critical damages are left behind on the track since this is mostly an ocular inspection done by of human.

Studies have also shown that imperfections in wheels cause an increase in the energy required to haul a vehicle and the noise generated by the vehicle [2, 1, 8]. They also contribute to escalated track dynamics and cause overloading on the track which reduces the remaining useful life and forces re-investment to be done earlier then the planned.

Due to safety and performance requirements, wheel-rail interface is a very important part, thus it is necessary to have a good control and monitor over the interface. The wheel and rail are closely dependent on each other and one has to consider both of them. It is necessary that infrastructure manager takes this into account,

not only the rail but even the wheels on rail [15]. The condition monitoring of both infrastructure and rolling stock is close related and the infrastructure manager need a good knowledge of both of them and treat this as a system. Therefore this paper investigates the monitoring systems of both the inspections of rails and wheels in a holistic view. However the problem of the wheel flat and other wheel imperfection is not only limited to direct economic and safety consequence. It also have additional operational consequence of extra track possession time and capacity reduction. Thus, it is herein referred to as failure driven capacity consuming events on the track.

There are some works already done to combine different way-side measurement equipment data but this is still at the developing stage and there is more to be done. Fröhling [5] concluded his paper with following sentence *“Much is, however still to be done before the wheel-rail interference system can be managed at a level in line with the available measuring technologies.”*

This paper presents different monitoring systems for railway infrastructure both the wayside and on-board systems. A case study of wheel imperfection detector and some measurements from the wheel profile measurement equipment are presented as well. The existing e-maintenance solution and the information flow of alarms. The paper is concluded with some discussions on how the information from these two equipment can be combined for a more holistic view, to improve the maintenance decision process for the infrastructure manager.

2. PARADIGM SHIFT IN MAINTENANCE

Maintenance is no more only a problem for the company, as it was before, it is an opportunity and value added process to have high safety, good available and sustainable process. The need to create additional value in maintenance and linking failure mechanism to life cycle management is the driving force for implementing prognostic and health management (PHM) of wheel condition. Figure 2 shows the connection between the general paradigm shift in maintenance and development of wheel condition monitoring systems over the years [9].

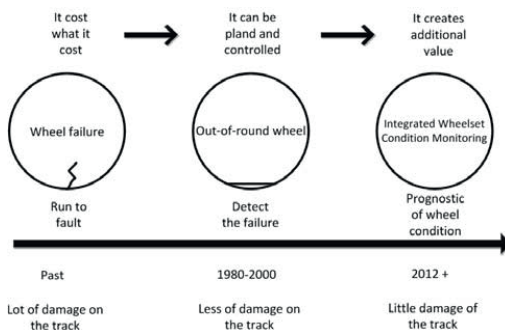


Figure 2. The paradigm shift in the wheel management, adapted from [9].

In the paradigm shift, for instance, the condition monitoring plays a key role. As it has the capability to capture condition data, that can be translated into information and the information can be

converted to knowledge by the e-maintenance solution, to enhance the planning of maintenance in an appropriate way.

This new knowledge will help to shift maintenance from routine-functional approach to process oriented approach with lot of effort in root cause elimination. From reactive to more proactive and business focused maintenance.

When the view of maintenance process is value added process, finding and elimination of root cause becomes an important concern. This concept is facilitated by condition monitoring and appropriate e-maintenance intelligence.

In the paradigm shift of condition monitoring in railway, the focus has shifted from detection to prognosis. Now the problem is not the data from monitoring equipment but the challenge is the interpretation of the data and combinations of data from different measurement equipment to deliver the right information in right time to the stakeholders.

One obvious implication of the paradigm shift in operation and maintenance of railway infrastructure in Sweden is the appreciation of safety. The focus on safety has increased and the reflection is on the trends of derailments. Number of derailments of travelling train on Swedish network from 1991 to 2011 have decreased from 46 to 7, see Figure 3 for detail [17].

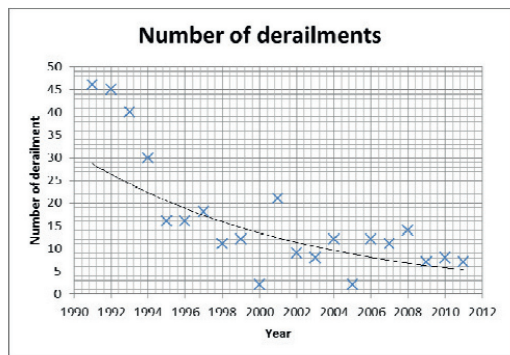


Figure 3. Number of derailments with train in motion on the Swedish railway network between 1991 and 2011 [17].

The safety on the railway can be further improved by paying more attentions on condition monitoring of the infrastructure and the rolling stock [4, 13]. This can minimise the “bad actors” on the track and focus on the problem areas for the infrastructure. A good status of the assets and rolling stock is a key issue to prevent accidents and failures, since accidents and derailments often causes of bad status of the equipment.

3. CONDITION MONITORING SYSTEM

In the railway transport system different condition monitoring systems are in existence. They are often divided into two: the on-board and wayside systems. The on-board systems measure the track and the wayside systems measure the rolling stock. Another classification based on the approach to maintenance is reactive or proactive system which has been mentioned earlier [10].

The Swedish Transport Administration, Trafikverket uses a lot of condition monitoring systems for both the track and the rolling stock. Some of these are presented below

3.1 On-board Inspection System

The measurement of the track is done by on-board inspection system (OIS) for the railway network in Sweden. The inspection class of the track decides how often the track will be inspected for safety reason. The network is divided into five different inspection classes B1 to B5 where B5 is the highest frequency class. These different inspection classes depend on the speed of the track and the yearly gross tonnage see Figure 4.

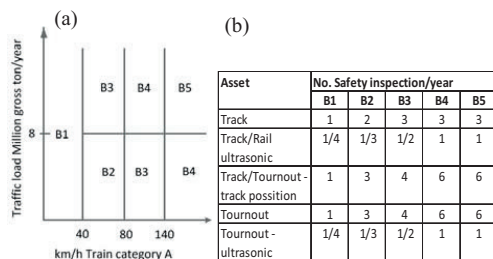


Figure 4. a) Inspection classes for the Swedish railway b) Inspection frequency for the classes of Swedish railway [18].

The inspection classes decide the number of safety inspections every year for each on board monitoring technique. For instance the safety inspections for the track in inspection class B5 have to be done three times each year see Figure 4b.

The track geometry quality measurement is done by the contractor. These cars measure for instance track quality, rail wear, ballast profile and the catenary wire, there is also video recording for the surrounding and the catenary. A single maintenance service provider does the track geometry measurement in the whole Swedish railway network.

The ultrasonic inspection is done by another contractor, and the inspection frequency also depends on the speed and track load, see Figure 4a-b. The eddy-current measurement is done only on special track with for instance lot of traffic and large axle loads. For the iron ore line these measurements have been carried out since 2006, before and after the grinding campaign every year.

3.2 Wayside Inspection Device

There are numerous wayside inspection devices (WID) installed along the Swedish railway network. To inspect the rolling stock there are wayside detector stations for the detection of hot boxes, hot/cold wheels, damaged wheels, overloaded cars, unbalanced loads, contact wire lift, pantograph and wheel-rail forces [16]. On the Swedish iron ore line, between Norwegian border and the harbour in Luleå there are sixteen WID installed, see Figure 5. Each point on the figure is a WID, and the exact location is written in the text. The wheel defect detectors on two track sections of the iron ore line is described below. One of these sections is located on the northern part and the other is on the southern part of the iron ore line. These two WID locations cover the whole iron ore line.



Figure 5. Wayside monitoring systems for the iron ore line in Northern Sweden [18].

An additional WID recently installed on the iron ore line, is the wheel profile measurement (WPM) equipment. It is the first of its kind in Northern Europe used for measuring wheel profiles on the regular line and captures the wheels of train with speeds up to 130 km/h. The installation and the operation is a collaboration in-between the main operator on the line and the infrastructure manager. This equipment has been in use for one year and has already done assessed to have a good performance up till now. Wheel with bad profile have been replaced after alarms.

3.3 Wheel Defect Detectors

The wheel defect detectors (WDD) use strain gauges, accelerometers or optical sensors to measure the wheels of passing trains [4]. The WDD measure the force from the wheel. In Figure 6, the straight line is the load on the loaded or unloaded wheel; the force peak comes from wheel defect [1]. The definitions of the forces are; F_p force peak, F_{dyn} dynamic supplement, F_l the wheel load from train and R is the ratio between F_p and F_l , see Formula 1. The force peak and the noise level increase with the train speed but on the other hand, the axle load of the train doesn't influence the noise due to roughness so much [11].

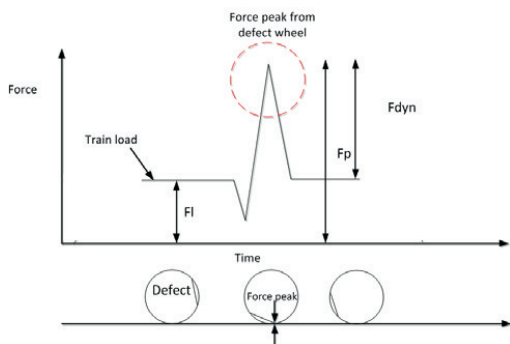


Figure 6. Hypothetical description of force load of a wheel defect, e adapted from [8].

$$Ratio (R) = \frac{F_p}{F_l} \quad (1)$$

Normal alarm limits for this forces are, force peak from 240kN up to 425kN, the dynamic supplement is in-between 160-240kN and

the ratio limit is between 3,7-4,2 this values depend on the train type.

The failure at the wheel is not every time easy to see visually, only the half of all wheels with force peak over 400kN has visual damage that was unacceptable [12].

The wheel defect detectors (WDD) on the Swedish network is in wide use and had since 1996 delivered information on bad wheels to the infrastructure manager and the rolling stock companies. The number of locations where WDD are installed is twelve; two of them are on iron ore line. Wheel defects are for instance referred to as out-of-round wheels and an example is wheel flat and below is a short introduction of both.

3.3.1 Out-of-round wheels

An out-of-round (OOR) wheel is a wheel with some kind of deformation on the surface; this deformation can have lot of shapes and have different root causes. This wheel defect causes dynamical forces damage to the rolling stock and the track. The development of the irregularities on the wheels depends on the dynamics of the system- rolling stock and the track [12,2]. The classification of OOR can be done as follow, eccentricity, discrete defects, corrugation, periodic non-roundness, non-periodic non-roundness, roughness, flats, spalling and shelling [12]. These defects can even be divided to two types, Type A - Tread Defects Initiated and Type B - Polygonization [19].

3.3.2 Wheel flats (skid flats)

A wheel flat is a flat part on the wheel, there are many causes of wheel flat; locking brakes, bad conditions of brakes, frozen brakes, bad adhesion in between the wheel and rail for instance due to leaves on the rail.

A wheel flat is classed as OOR Type A defect [19]. Typical parametric description of wheel flat include length and height of the defect, these are shown in Figure 7.

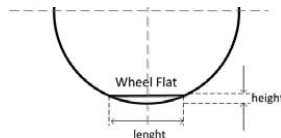


Figure 7. Flat wheel, the length and the height of a flat.

The approximate length of a flat for an alarm is around 60 mm [11]

Furthermore, studies have shown that flat grows rapidly in the beginning and on afterwards with sliding the growth decreases. It is usual that cracks are initiated on wheel flat after sliding; research has shown that almost 66% of all the wheels after the sliding have cracks [8].

3.4 Wheel profile measurement

In the market a lot of different wheel profile measurement (WPM) systems can be found, for instance Brickle et al. [4] has reported at least twelve of the WPM equipment presently available in the open market. All these systems differ in technology, configuration, precision, winter reliability, capacity and other characteristic features that can be used for selection purpose. Some of the common wheel profile measurement equipment

available in market include: Wheelview, WheelScan, FactIS WPM, Treadview and Argus [4].

A wheel profile is not constant around the whole wheel; there is an average variation for the circumferential for the flange height within 0,131 mm, for the flange thickness 0,145 mm and for the hollow wear 0,087 mm [5]. This variation has to be taken into count also the accuracy of the measurement equipment.

The only Swedish wheel profile measurement equipment is located on the track section which lies on the southernmost part of the iron ore line. This system has been in use since autumn 2011 and the data from the initial run was delivered to the operator but in the future it will be delivered to the infrastructure manager. The installation of the system was collaboration between the infrastructure manager and the main operator on this line. This system will deliver necessary information to the infrastructure manager about the wheel fleet on the track. The information will help the infrastructure manager to build up the maintenance strategy and give opportunity to develop the prediction of the life length for the track in the future. This information can be used in track Life Cycle Cost (LCC) calculations as an input to estimate the rail degradation.

The wheel profile measurement equipment consists of four units, two on gauge sides and two on the field sides of the rails. These units contain a laser, a high-speed camera, and an electronic control system. When a train passes the boxes, the first wheel triggers a sensor and the protection cover opens, the laser beam starts to shine, and then the camera takes pictures of the laser beam projected onto the surface of the passing wheels. These pictures are saved and an algorithm transfers the pictures of the wheel profiles to an "xy-coordinate system". The coordinates can be shown using software and can be compared to the nominal wheel profile. Figure 8 shows the wheel profiles from the measurement equipment, one profile from a locomotive and one from a wagon, both from one iron ore train in 2012-06-06. The x and y scale is in mm.

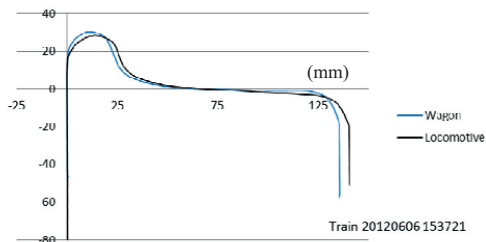


Figure 8. Two wheel profiles from iron ore train measured by the wheel profile measurement system in Sunderbyn.

This system measure for instance; flange height, flange width, flange slope, tread hollow and rim thickness up to a speed of 130 km/h.

After the first year in use, the system given fairly good results but there are potential to improve the reliability of the system. There are some operational problems that have to be resolved for the next season for reliability improvement.

The system is in operation and has already captured and alarm wheels that are beyond the safety limits. This information is used for maintenance action of re-profiling the wheels.

The reasons for re-profiling the wheels are majorly wear of the flange or the hollow wear. The re-profiling due to the flange wear usually need larger cuts than the hollow wheel wear, this mean more material have to be removed with the flange wheel wear [5]. Number of re-profiling increases and the dominant failure mode are RCF (spalling and shelling) in lower operating temperature [14].

4. CASE STUDY

This chapter shows a case study of wheel defects from the Swedish railway on the two track sections on the Swedish iron line describe earlier. Wheel defects is a capacity consumption failure mode due to at least two aspects; the first is the direct stop in traffic due to the speed limitation after a detection of defect wheels, the next is abnormal deterioration of the rail, fastener and other track components. The data analysed is for and includes 1509 wheel defect incidents. The incidents include three levels; warning, low alarm and high alarm. The total number of trains passing both detectors for year 2011 is around 24000.

4.1 Data analyses

Figure 9 shows all events divided into 13 categories with 5 degree interval from -32,5°C to +32,5°C. This implies that k intervals of $(s_0, s_1], (s_1, s_2], \dots (s_{k-1}, s_k]$ can be obtained where $s_0 = -32,5^\circ\text{C}$, $s_k = 32,5^\circ\text{C}$ and $k=13$. It can be seen from the Figure that there are so many wheel defects records during the colder season with temperatures under 0°C, but the propensity of the phenomena is reduced in warmer temperature.

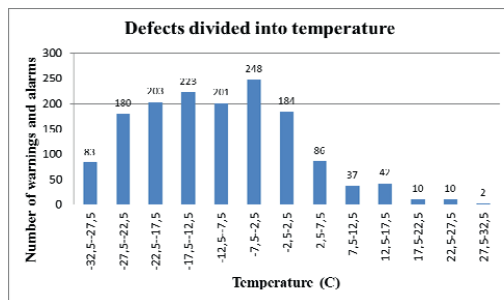


Figure 9. Number of wheel defects divided into temperature spans of 5 degrees.

Another notable observation in the wheel defects records is the extreme wheel flat alarm condition. That is to say there are some train passages with many warnings and alarms from the WDD for wheels on same train and single passage. Trains with most number of wheel flats are presented in Figure 10. The maximum number of warnings and alarms for defect wheels for one train passage is seventeen. Additional information in this figure is the running direction of the train, which is either odd or even direction. The even direction is in direction to the harbour with loaded 68 iron ore wagons having at least 30 tonne axle load and

the odd direction is direction from the harbour with empty wagons with lower axle load. The air temperature during the actual detection time of the high alarm trains is shown by the line plot in Figure 10. The, average air temperature for all this high alarm wheel defects incidence is -17.3°C . Nonetheless wheel flat has also been recorded at high temperature, showing that this has occurred over a large temperature range- 60°C for the air and 71°C for the rail in the case study

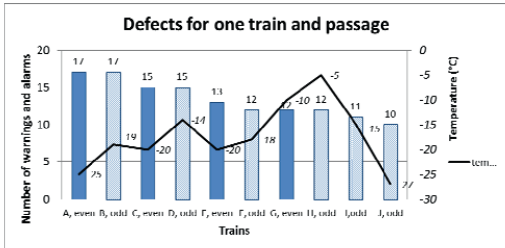


Figure 10. Number of warnings and alarms of wheel defects for single passage of selected trains with respective air temperature on the site.

Considering only the high warning and alarm trains, recorded accumulated incidence in both even and odd direction differ. In the even direction there are 524 warning and alarms, in odd direction there are 985 see Figure 11. To establish the underlying explanation for this effect, further investigation is required. This directional effect could be due to many reasons among which are gradients, loading condition, winter effect (less snow towards the harbour than far away from the harbour). On the other side it might be mere coincidence for the year and data being studied.

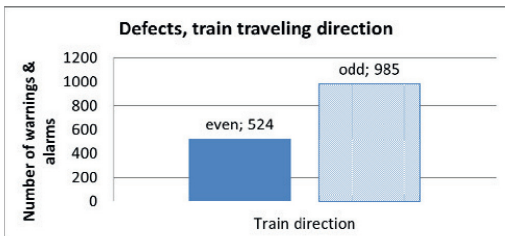


Figure 11. Train direction for train with wheel defects.

The wheel position and number of warnings and alarms for wheel defects of the freight train fleet for year 2011 is shown in Figure 12 and 13. The freight trains are divided in to two groups, locomotives with three axles in each bogie (iron ore trains) and locomotives with two axles in each bogie (other freight trains). Figure 12 shows the trains with three axle bogies. 6,5% and 5% of all wheel defect warnings and alarms are from axles six and seven.

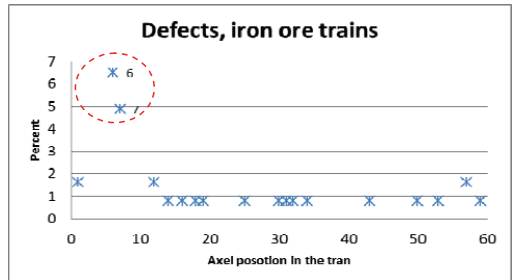


Figure 12. Wheel defect warning and alarms of freight train with three axles in each bogie.

Figure 13 shows the train with two axles bogies, here there are no distinct axles with high wheel defect warning and alarms. Nonetheless axle positions nine and twenty-nine have higher records of warnings and alarms- 2,5% and 2.3% respectively of all alarms and warnings, for this train types.

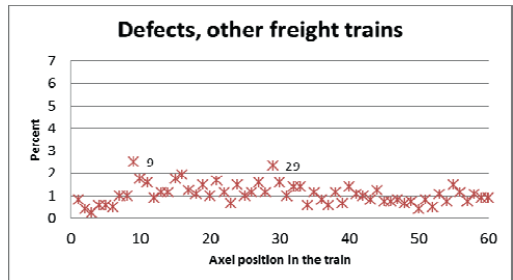


Figure 13. Wheel defect warnings and alarms from freight trains with two axles in each bogie.

As mentioned earlier, wheel flat and other out-of-round defects shoot up the dynamic supplement forces, from the wheel, in this case up to 459kN while the lowest alarm limit is 240kN. The study of the wheel flats in 2011 on the mentioned track sections gives the following information. The air temperature appears to have influence on the number of wheel defects. Low temperature increases the chance of wheel defects occurrence. Number of detected wheel defects on one train varies from one to seventeen wheel warnings and alarms. The transport direction has an influence on the number of warnings and alarms of wheel defects. Then there are wheel positions with more number of warnings and alarms of wheel defects than the other.

5. COMBINATION OF WHEEL DEFECT DETECTOR AND WHEEL PROFILE MEASUREMENT

The combination of the data from the WPM equipment and the wheel defect detector will increase the information and knowledge of the condition for rolling stocks and a more holistic view of the status on the wheel. For instance, if the status of a wheel passing through the profile measurement equipment is bad but has not exceeded the alarm limit, and the same wheel has even some out-of-round defects under the alarm limit, this gives two

different information. This information together can show a totally new picture of the wheel status. These two defects together can boost up the impact of the wheel on the track more than a wheel with perhaps an alarm from only one system. The knowledge can be used as basis for decision making by the infrastructure manager but also for the rolling stock operator. This knowledge will be facilitated by a good e-maintenance solution- a useful tool for integration and management of data from different sources for effective maintenance decisions and actions.

The overlaid information from both systems can together give new possibility to find and isolated defected wheels in an early stage to prevent serious failures. Prediction of the rail degradation process and preservation of the quality of both rail and wheel will be enhanced by the combination of the information from the two monitoring equipment. This will even contribute towards safety improvement as it will point out the "bad actors" in a reliable way, removed then from the track in a proactive approach before any damage will be done.

The combination of this system can increase the advantages of the condition monitoring and might even the presents some challenges.

Some advantages are, improving of safety, reliability and efficiency, increased train availability, more consistent inspections, reduced number of train delays, prioritised and focused maintenance, extended vehicle service life, historical database for trending and eventual predictive maintenance [4, 13]. Also the combination possibility can be of advantage, to define new more dynamical thresholds' when information are from different system are integrated.

There are even some disadvantages that have to be considered; system maintenance cost, potential for too much data and increased complexity [4, 13].

5.1 e-maintenance concept for wheel condition monitoring

The information aspect of PHM of wheel condition is very essential for the success of the approach. The centre point of PHM is to use information to make early detection of impending or potential failure, facilitate remaining useful life calculations, and support effective decision-making based on predictions. In physical asset management, common approaches for implementing prognostics and health monitoring are (i) installing built-in/external units that will fail/trigger when certain operating condition is exceeded- e.g. wheel defect detector; (ii) monitoring environmental and usage data that influence the system's health and converting the measured data into life consumed; (iii) monitoring parameters of system conditions which can reflect potential failure - e.g., system characteristics, defects, performance [20]. In the wheel condition monitoring described in this study, all the three approaches mentioned above are relevant. Integrating the data and information from these approaches requires a well-designed e-maintenance framework. The requirements, expectation and performance of the e-maintenance solution as related to effective infrastructure maintenance through PHM of wheel condition are described below.

The e-maintenance framework to support the PHM of wheel and rail condition is shown in Figure 14. It shows how the information from lot of sources, for instance wheel defect detector and wheel

profile measurement system, can be managed to benefit both rolling stock operator and the infrastructure manager.

The lower part of the figure (under the dotted line) shows the existing framework of how the alarm works today; this is more relevant to traffic operation. In this framework there is no integration of the information from the different detection systems, only one system can give alarm at a time when a threshold is reached. Then the wheel profile measurement system is not integrated into the condition monitoring systems that is in use today for the Swedish network.

The upper part of the figure is the additional e-maintenance solution; this gives possibilities to combine the data from different measurement equipment and apply dynamical thresholds, deliver prognostics and diagnostics information to the operator and the infrastructure manager, same picture of the status but the information can be adapted to the needs. This information goes to the operator and to the infrastructure manager's maintenance organisation and serve as basis for planning and scheduling of maintenance activities.

The information to the database comes from many different data sources that include track data, train data, time tables and lot more. The condition monitoring equipment gives also information to the data base. The data from the database goes to the Traffic Control Centre (TCC) and to the e-maintenance system.

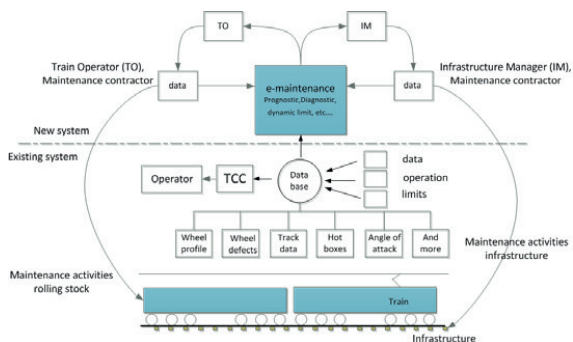


Figure 14. e-maintenance solution for Prognostic and Health Management (PHM) of wheel condition

The e-maintenance supports diagnostic, prognostic, and dynamic limits for wheels. All this are built on the information from the database and information from the maintenance organisation of the rolling stock and the infrastructure manager. And it will facilitate proactive maintenance planning and actions for both train operator and the infrastructure manager.

However, there are also issues and challenges with the e-maintenance concept in the figure above. There are lot of work to be done inside of the organization to support the framework after implementation. The database need to have capability and availability to, take care of all types of data, lot of data from the stations, merge the data together, report this data in a useful and easy way and then ensure the information is tailored to the need of the stakeholder [4].

6. CONCLUSIONS

There is a lot to gain with good condition monitoring of the rolling stock, both the infrastructure manager and the operator can benefit from this.

- The patterns of WDD differ between different trains. For the iron ore train's two axels can be specially pointed out with a higher frequency of wheel with large forces and probably more defects on the wheels.
- The temperature seems to have some influence on the numbers of defected wheels, lower temperature contribute more to development of wheel defects.
- The directions of the train have some influence on the number of defected wheels, the direction from harbour have more defect wheels.
- A combination of WPM and WDD can give additional information on wheels that degrade the track, but not exceed the threshold limits for each system. This combination gives possibilities to develop dynamic thresholds for wheels on track. The information of the status can come from many systems and can be combined with track data.

The paradigm shift in the condition monitoring is essential to improve railway infrastructure maintenance so as to support the extension of the remaining useful life (RUL) for rolling stock and infrastructure.

6.1 Further work

The study suggests the following additional research:

- Find the thresholds for the combination of the data from the WDD and WPM systems.
- Investigate the differences between the wheel defect patterns of the different freight trains.

ACKNOWLEDGMENTS

The authors wish to thank Trafikverket (the Swedish Transport Administration) for the data to this paper and Luleå Railway Research Center (JVTC) for financing this work.

REFERENCES

- [1] Baeza, L. Roda, A. Carballeira, J. and Giner, E. (2006). "Railway train-track dynamics for wheel flats with improved contact models," *Nonlinear Dyn.*, vol. 45, no. 3-4, pp. 385.
- [2] Barke, D. W. Chiu W. K. (2005). "A review of the effects of out-of-round wheels on track and vehicle components," *Proceedings of the Institution of Mechanical Engineers.Part F, Journal of rail and rapid transit*, vol. 219, no. 3, pp. 151.
- [3] Barke, D. Chiu, W. K. (2005). "Structural health monitoring in the railway industry: A review," *Structural health monitoring*, vol. 4, no. 1, pp. 81.
- [4] Brickle, B. Morgan, R. Smith, E. Brosseau, J. and Pinney, C. (2008). *Identification of Existing and New Technologies for Wheelset Condition Monitoring: RSSB Report for Task T607, TTCI (UK) Ltd.* Rail Safety and Standards Board .
- [5] Fröhling, R. D. and Hettasch, G. (2010). "Wheel-rail interface management: A rolling stock perspective,"

- Proceedings of the Institution of Mechanical Engineers. Part F, Journal of rail and rapid transit*, vol. 224, no. 5, pp. 491.
- [6] Jardine A. K. S. and Tsang, A.H.C. (2006). *Maintenance, Replacement, and Reliability : Theory and Applications*. Florida, CRC Press, Taylor and Francis Group.
 - [7] Jämvägsstyrelsen (2008). *Jämvägsstyrelsens trafikföreskrifter, JvSFS 2008:7, bilaga 6*, <http://www.transportstyrelsen.se/Regler/Regler-for-jarnvag/JTF/2008>.
 - [8] Jergéus, J. (1999) "Full-scale railway wheel flat experiments," *Proceedings of the Institution of Mechanical Engineers.Part F, Journal of rail and rapid transit*, vol. 213, no. 1, pp. 1.
 - [9] Kumar, U. (2008). "System maintenance: trends in management and technology," in *Handbook on Performability Engineering*, K.B. Misra, Ed. London: Springer, pp. 773-787.
 - [10] Lagneback, R. (2007). *Evaluation of wayside condition monitoring technologies for condition-based maintenance of railway vehicles*, Luleå University of Technology, Sweden .
 - [11] Madejski, J. (2006) "Automatic detection of flats on the rolling stock wheels," *Journal of achievements in materials and manufacturing engineering*, vol. 16, no. 1-2, pp. 160.
 - [12] Nielsen, J. C. O. Johansson, A. (2000). "Out-of-round railway wheels-a literature survey," *Proceedings of the Institution of Mechanical Engineers.Part F, Journal of rail and rapid transit*, vol. 214, no. 2, pp. 79.
 - [13] Ollier, D.B. (2006) "Intelligent infrastructure-the business challenge," vol. 2006, pp. 1 2006.
 - [14] Palo, M. (2012) *Condition monitoring of railway vehicles*, division of operations, maintenance and acoustics, Luleå University of Technology, Sweden.
 - [15] Smith, R. A. (2003). "The wheel-rail interface—some recent accidents," *Fatigue fracture of engineering materials structures*, vol. 26, no. 10, pp. 901.
 - [16] Stenström, C. Parida A. and Galar, D. (2012). "Maintenance performance indicators for railway infrastructure management," in *Proceedings of the first international conference on railway technology: Research, development and maintenance*, 2012.
 - [17] Trafikanalys, (2012). *Rail traffic accidents 2011*, Trafikanalys, www.trafa.se.
 - [18] Trafikverket, (2005). *Safety inspections for permanent assets (Säkerhetsbesiktning av fasta anläggningar, BVF 807.2)*, Trafikverket, F04-3712/UN20 2005.
 - [19] UIC (2004). "Atlas of Wheel and Rail Defects," *INTERNATIONAL UNION OF RAILWAYS*, pp 36-37.
 - [20] Vichare, N. Tuchband, B. and Pecht, M. (2008). "Prognostics and Health Monitoring of Electronics," in *Handbook of performability Engineering*, K.B. Misra, Ed: Springer, pp. 1107-1123.

Improvement of configuration management in railway signalling system

Amparo Morant

Ramin Karim

Phillip Tretten

Diego Galar

Operation and Maintenance Engineering
Luleå University of technology
Luleå, 97 817, Sweden

+46 (0)920 4925188
amparo.morant@ltu.se

+46 (0) 920 492344
ramin.karim@ltu.se

+46 (0) 920 492855
phillip.tretten@ltu.se

+46 (0) 920 492437
diego.galar@ltu.se

ABSTRACT

The railway network composed by different systems, such as signalling, power supply, rolling stock and others. The signalling system is used for control, supervision and protection of railway traffic. Its reliability, maintainability and related maintenance support affect the availability of the railway network. Often, signalling system itself is considered as a system-of-systems spread over a wide geographical area and consists of a large number of items with different lifecycle. However, managing a complex technical system with large number of items in different lifecycle phases to achieve required availability is challenging. One major challenge is to define and describe the structure and relation between the system's inherent items at any specific period of time, i.e. the configuration of the system. The system configuration is important during the design and manufacturing of systems, but it is also essential during the utilisation and retirement of a system. Furthermore, system configuration can be considered as an essential data container to which other data sources can be related, e.g. design data, reliability data, maintainability data, operation data, and maintenance data. A proper management of configuration data is highly important during a system's whole lifecycle, e.g. the railway signalling system. Hence, the purpose of this paper is investigate how the process of configuration management related to railway signalling system can be improved through utilisation of Capability Maturity Model Integrated (CMMI).

Keywords

Railway, maintenance, configuration management, signalling systems, CMMI,

1. INTRODUCTION

In railway, the signalling system plays an important role in the control, supervision and protection of the traffic. Hence, its availability performance is directly related to and affects the overall performance of the whole system.

The railway network can be divided in different systems depending on their functionality, such as the rolling stock, the track, the power supply, the signalling system, etc. [31]. For the signalling systems different solutions can be found [1]. Track circuits, axle counters and GPS-based systems can be used to locate a train. Track circuits and signals can also help to control the traffic on the railway line to prevent collision. Balises and radio based systems allow the train control centre to restrict the movement of trains, and advanced systems i.e. European Rail Traffic Management System (ERTMS) or Automatic train control system (ATC) provide the supervision and control of the railway network. They are in charge of the interpretation of the inputs

from the other systems, creating restrictions on the train route to ensure a safe operation. Today, there are many signalling systems on the market, each with different specifications and based on different technologies (e.g. ERTMS, ATC, ATS, ASFA, etc.) [1, 15, 28, 37]. A summary of the different parts on a signalling system are shown in Figure 1.

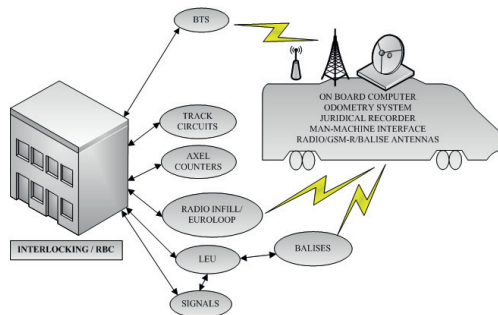


Figure 1. Different signalling subsystems and their interfaces.

The main characteristic of a signalling system is that the overall function is fulfilled by the sum of the functionalities of the different parts: the supervision and protection of the railway network will not be possible if any of the items of the signalling system don't work properly, each part has an independent function to perform. These characteristics are the ones that define a system of systems (SoS) [3]. When managing SoS it is not possible to consider the different parts independently, since the relation between the different item is also important to achieve the desired functionalities [5].

The origin of signalling is closely linked to the development of railways. From the initial manual restrictions to access a line, the growing demand for transportation, and the increasing number of trains running, made this system not enough. Advanced technologies were implemented in order to perform the supervision and the control of the railway lines. These systems were mainly based in analogue systems, based on relay technology (e.g. track circuit, axle counter, relay interlocking). Nowadays these systems are being replaced for digital control systems based on electronics (e.g. balise, electronic interlocking, lineside electronic unit – LEU), but still both systems coexist in most of the railway network. Over the years, many signalling and train control systems have evolved, creating a highly technical and complex industry. Nowadays there is a need to be able to operate the trains not depending on the country. That made the need to

have a unified signalling system that could avoid the changes needed nowadays for going through different countries. ERTMS was the solution for that. In order to achieve interoperability between the control and supervision systems, several contributions via standardisation have been provided [35, 36]. Standards have been developed also regarding the reliability, availability, maintainability and safety (RAMS) of the different railway systems [6], with special focus on the systems regarding the signalling, communications and processing systems on the railway [7, 8, 9]. These standards aim to enable interoperability of the line without affecting the safety of the system.

Signalling systems comprise a wide range of technologies and equipment. In some countries (e.g. Spain) the following signalling systems are used: LZB (Linienzugbeeinflussung - Continuous train control), ASFA (Anuncio de Señales y Frenado Automático - Announcement of signals and automatic break), ERTMS, and several versions of ERTMS (2.3.0d, 2.3.0, 2.2.2+) [19]. The infrastructure is provided by Alstom, Thales, Invensys and Ansaldo; the onboard signalling systems come from Bombardier, Alstom, Ansaldo, Invensys and Siemens. Meanwhile, Sweden uses the signalling systems ATC and ERTMS from Bombardier and Ansaldo and an onboard signalling system from Bombardier.

From a maintenance perspective the railway signalling system should include both functional and structural aspects. Most of the signalling systems today consist of electronic items, in which software is critical. Furthermore, it is essential to control the configurations and changes in the software installed in the various system elements through the whole lifecycle of a system. Hence, the structural aspects should include all items included in the system, both software and hardware.

There are different challenges when performing maintenance on a SoS that usually don't exist on a system. Interoperability should be ensured and, for that, it is necessary that the different configurations of each item are compatible between them. A good maintenance performance will take these compatibilities into account, since it is necessary to control the information of which constitutes the signalling system and how the signalling system is affected by any changed performance. However, managing configuration information in such a heterogeneous context is complex and requires an appropriate process for configuration management [2].

Extensive application of ICT (Information and Communication Technology) and other emerging technologies facilitate easy and effective collection of data and information [29]. In maintenance, enhanced use of ICT facilitates the development of artifacts (e.g. frameworks, tools, methodologies and technologies); which aim to support maintenance decision-making. These artefacts also enable improvement of different maintenance approaches, such as; preventive maintenance and corrective maintenance. Furthermore, ICT provide additional capabilities, which can be used within diagnostic and prognostic processes. The prognostic and diagnostic processes in an enterprise can be facilitated through provision of proper information logistics aimed to support maintenance decision making through provision ICT-based solution for data processing and information management [23]. In the context of complex technical systems information logistics refers to: 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'; 4)

context management, which addresses 'where and why to deliver' [22, 23, 24].

IEV 394-33-27 defines configuration management (CM) as the process of identifying and documenting the characteristics of a facility's structures, systems and components (including computer systems and software), and of ensuring that changes to these characteristics are properly developed, assessed, approved, issued, implemented, verified, recorded and incorporated into the facility documentation [18].

Configuration management is also taken into account within the railway standards. The standard EN 50126 defines the process of configuration management (CM) as the discipline that, applying technical and administrative direction and surveillance, identifies and documents the functional and physical characteristics of a configuration item, control change to those characteristics, record and report change processing and implementation status and verify compliance with specified requirements [6].

The railway standard EN 50126 also defines and establishes also the relationship between RAMS (Reliability, availability, maintainability and safety) and Configuration Management. This standard defines RAMS in terms of reliability, availability, maintainability and safety and their interaction and defines a process, based on the system lifecycle and tasks within it, for managing RAMS; but also establishes some mandatory requirements. One of them is to establish and implement an adequate and effective configuration management, addressing RAMS tasks within all lifecycle phases.

The scope of the configuration management will depend on the system under consideration, but it shall normally include all system documentation and all other system deliverables. Configuration management processes are implemented usually in the design process, but not long after the system acceptance and during the operation and maintenance phase of the cycle of life. While in the design phases the main focus on the CM is to give traceability of the design changes and improvements in order to validate the final solution, during the operation and maintenance phase the scope in CM should be to give visibility of the entire system in order to keep the interoperability between the different parts. To trace any change performed on the installed equipment is also important and allows to keep control of the state of the items (e. g. how many devices are on repair).

When performing the maintenance of a system, which is a combination of software and hardware, the maintainer must understand how changes will affect the system, how the system is built and what role the different parts play and how they are interconnected. If up-to-date documentation is lacking, serious problems for the maintainers arise. [14]. The management of this information is usually performed by the CM process.

The infrastructure manager must ensure that the entire network fulfils the requirements (e.g. safety, functionality, and non-functionality) from various stakeholders (e.g. owners, suppliers, operators, customers, government, entrepreneurs, etc.). That can be ensured performing a good control of the information and any changed performed on the configuration of the network. In addition to that, the configuration needs to be managed and adapted to various stakeholders' processes (e.g. design, manufacturing, operation, and maintenance processes). The complexity of the signalling systems makes this CM a need in order to ensure a proper performance.

The CM process can be established or improved through various approaches such as Software process improvement (SPICE), Capability Maturity Model Integrated (CMMI), Six Sigma, Total Quality Management (TQM), etc. Some of them are general methods to improve processes (Six Sigma), others are focused on software management (SPICE), and others establish a model but not the specific process (CMMI). In particular, CM on signalling systems shall take into account the approaches that allow the configuration management of both software and hardware in the same process model.

The software engineering institute (SEI) has developed CMMI (Capability Maturity Model Integrated). CMMI is a framework that helps organizations to improve their processes [4]. CMMI was generated from previous models, including software management models (like the CMM model), and nowadays it includes processes of software and hardware management which makes it a good choice when performing CM in signalling systems.

Hence, the purpose of this paper is to investigate how the process of configuration management, related to railway signalling system, can be improved through utilisation of the CMMI model.

2. RESEARCH METHODOLOGY

Little has been done on CM for signalling systems on the operation and maintenance phase of the cycle of life, however some research has been done in related areas, such as CM in software maintenance, CMMI based models on signalling systems, etc. In our case, the work is based on a study of the state of the art on CM and different surveys in the companies in order to define the requirements to fulfil by the CM process.

Regarding the industry, several interviews were made in order to know the needs of the different stakeholders of the CM process, such as system engineers, maintenance managers, maintenance personnel, quality and RAMS managers, etc.

In these interviews the information to control was identified, such as procedures and manuals, mechanical drawings and electrical schemas, software files, hardware serial numbers, etc. It was also defined which information was not needed to control, e.g. fasteners, cables and internal documentation.

From the different surveys and the research literature some conclusions can be underlined. Maintenance management on signalling systems has a challenge on the amount of information needed to perform a good preventive and corrective maintenance. Not to have the proper data can lead on an increase time on corrective maintenance actions due to a wrong failure identification, which brings a lower availability of the system. Even when the failure is well identified, if the item is not repaired assuring the interoperability with the rest of the systems, the system will not be operable. Hence, a configuration management process is a must.

The main purpose of the process developed is to give a tool that makes the accessibility of the information and its control easier to all the people involved on the signalling systems maintenance and operation. In other words, to implement CM as a useful tool for the people involved on the operation and maintenance of the signalling system.

After studying the different possibilities to improve the CM process, we concluded CMMI can be one of the best approaches to solve the challenges on signalling systems. CMMI is a model,

not a rigid method, hence allows adapting the model to the requirements of the system. Other methods like KM or Six Sigma are more focused in process improvement; therefore they can be combined with CMMI to improve processes. The Agile method is specialised on software management so does not fulfil the requirement of manage the configuration of both hardware and software.

3. STATE OF THE ART

Some publications can be found regarding CM on maintenance that can be extrapolated to railway signalling systems. Kitahara et al proposed a method to perform CM on the maintenance of the railway software, but do not take into account the control of the hardware [26]. Kelly and McDermid exposed a change control management process for the safety case during the maintenance, with an example on the aviation sector that can be applied on the railway sector [25]. Pasquale et al proposes a process methodology to perform the hazard analysis on complex distributed railway systems [30]. Joung and Shin described a methodology to develop railway software based on the CMMI model [21].

De Souza et al exposed the documentation needed to consider during the maintenance of a software based system as a result of a different surveys [14]. Liu exposed the problems on software maintenance and the improvements when implementing the CMMI model [27].

Turner and Jain studied the differences between different approaches for process management, such as the differences between the Agile approach and the CMMI model for software management [34]. Dayan and Evans exposed the similarities between the CMMI model approach and the Knowledge Management (KM) [13]. Sutherland discussed how to combine on software management Agile and CMMI approaches [33]. SEI has published some studies comparing the CMMI model with other approaches on process improvement, such as how to combine CMMI with Six Sigma [11] and why not to do combine CMMI with Agile [12].

Regarding CM process on SoS, some results have been published. Gorod et al describe the challenges of managing a SoS compared to a normal system, including the configuration management [17]. Bellomo and Smith II exposed the challenges of performing a configuration management process on a SoS [2].

Fonseca et al proposed a process following the CMMI model based on the CENELEC standards [16]. Jansson studied how the implementation of the CMMI model on the maintenance of software [20]. CMMI documents describe a CM management process divided on goals and tasks to perform [10]. A survey was made also to know which tools were used by the companies to apply the CM process. There are several tools in the market in order to perform the process of CM, depending on the requirements of the system. Ren performed an evaluation of some CM tools for software CM, such as CVS, Firefly, ClearCase and others [32]. The tool to implement our model will have to be agreed with the infrastructure manager depending on the needs and priorities.

Our case study is based at a company responsible for the management of the infrastructure. The configuration of the whole railway infrastructure is managed on a software tool which allows knowing which items compose a section of the railroad (signalling, power supply, track components, etc.). The specific location of each item is defined together with the model and the

serial number. This tool shows some information regarding corrective and preventive maintenance work performed on that item. There are also two software tools to manage both the corrective and preventive maintenance. The documentation is managed in a repository which can be accessed through internet. There is no relation between the system configuration and maintenance performance with the documentation needed to perform it, which can bring a poor visibility of how to maintain the system. The relations between the different systems are not visible, each item is described as independent, and therefore the interoperability after maintenance is not assured. It is not possible also to see which are the software and hardware tools involved in the maintenance tasks. Last but not least, the software of the different systems is not managed by the configuration tool.

4. CMMI: CAPABILITY MATURITY MODEL INTEGRATION

CMMI is a process improvement maturity model for the development of products and services [10]. Different adaptations depend on the company focus. CMMI for Development establishes a group of best practices for development and maintenance activities covering the life cycle of different systems, from conception to delivery and subsequent maintenance. ‘CMMI for Services’ is designed for businesses focusing on establishing, managing and delivering services. ‘CMMI for Acquisition’ is designed for businesses working with suppliers to assemble a product or deliver a service.

Depending on whether the focus is on an individual or multiple processes, the CMMI model establishes either capability or maturity levels, depending on the level that the company wants to achieve. Different processes are associated with each level. As shown in Figure 2, the levels depend on how well processes are implemented in the company, from the initial stage where there is no standard or homogeneous process, to the fifth level where all processes are implemented; at this level, they are homogeneous in all areas of the company and the focus is on their continuous improvement.

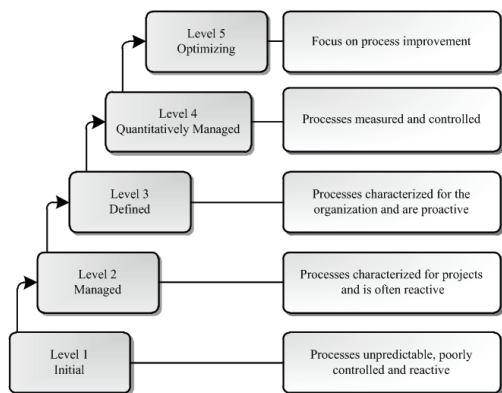


Figure 2. The maturity levels of the CMMI model.

In CMMI, a process area is a group of related practices that allows for a set of goals to be achieved, resulting with an important improvement in a specific area. The CMMI model establishes 24 process areas. Depending on the maturity level of the model, it

may be called upon to handle a differing number of process areas, as each maturity level represents a pre-defined group of process areas. For example, at level one (Initial), the processes are not controlled; as the level does not follow the CMMI Model, no processes are considered. At the second level of maturity (Managed) eight processes are involved; for the following levels, more processes are taken into account.

Signalling systems can be considered a system of systems (SoS). They are composed of different systems that have their own purpose but the main functionality is given by the interoperability between them. For optimal maintenance, it is necessary to keep track of the configuration of the different systems and where they are allocated. These configurations must consider not only the software and hardware of each subsystem, but also the interfaces between systems and where each device is located. Some examples of data requiring controlled configurations include the hardware and software of each device, the location of each device and the interfaces between systems.

One solution to this is to have a centralised approach with a common database that provides all the required data from the infrastructure to the various users and managers of the infrastructure. Since signalling systems now involve onboard systems, it would be possible to unify the infrastructure management with the management of the rolling stock.

A common database benefits all management companies, being able to share the common information and keep the rest of the data only visible for the company involved. This can be managed through development of a proper security system that manages accessibility to the system and information. For example, on a need to know basis, the operator manager could see the resources relevant for his specific needs.

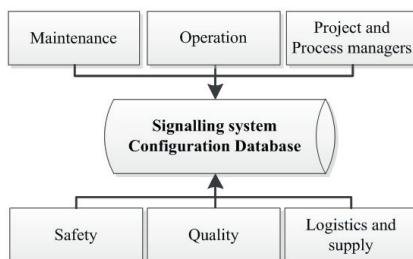


Figure 3. System configuration database users.

As shown in Figure 3, even within the same management company, different users require the information. This includes the following: to analyse corrective and predictive maintenance for maintenance managers, identification of tendencies and management improvements; maintenance crew can easily obtain information and material needed to perform repairs, as the work orders are linked to the configuration of the infrastructure; Logistics and supply managers can determine which devices are installed, which are being repaired and which are in stock; Operation managers can receive information on problems in the railway operation and determine their cause as well as, performing simulations of how a change in the configuration of the signalling system can affect the availability of the line and improve the capacity of the railway network; Quality and safety managers have easy access to the certificates of the devices and the

processes to perform any action on the infrastructure, thereby, reducing the time needed to do a review or audit or keep track of safety issues; Project and process managers have complete visibility of the management numbers obtained from the KPIs and, therefore, know the real situation of the infrastructure and processes.

5. IMPLEMENTATION OF CMMI IN RAILWAY SIGNALLING SYSTEM

The main tasks required for good performance of the configuration management process are the following : identify the configuration of the items selected in the system to control at either periodic or punctual points in time, depending on the system configuration; distribute periodic reports of the state of the processes and changes performed to provide information on the real status of the system to the different managers and end users; track the changes and modifications of both configuration and processes management by performing periodical reviews and audits, always looking forward to check the integrity of the configuration baselines and determine whether they reflect the real installation configuration status [10].

A description of the tasks and how they are applied in the configuration management process of a signalling system maintenance organization is given below, based on the requirements of the signalling system maintenance management and applying the CMMI model in order to improve the actual CM process.

5.1 Establish baselines

Specific practices and sub practices are defined for achieving the configuration management process based on the model. In the case of the specific practice of establishing the model, the sub practices include identifying the configuration items, establishing a configuration management system and creating or releasing baselines [10].

A baseline is the configuration of the system at a fixed point in time in which the configuration will be the reference for controlling any changes performed on the system. A baseline describes what items are part of the system and their status and it also provides documentation for and information on the real state of the system at that point of time. The concept of baseline could also be explained as if a picture of the system was taken at a particular time. Baselines describe the status of a determined system at a fixed point in time; they serve as a reference for tracking changes (such replacement of items due to failure) on that system. They can also be determined for a particular item and show the changes performed on it during a specific time.

Good configuration control of the information involved in signalling system maintenance requires that we consider the plethora of information considered in the baselines. Different baselines can be described depending on the configuration items considered. Signalling systems require documentation, design, and installation baselines. The documentation baseline gathers all data compiled on documentation, such as procedures, mechanical drawings, certificates, etc.

The system design baseline makes the structure of a system visible and indicates the models of the different subsystems and devices, together with the software that configures a specific signalling system. While, from the design structure, the installation baseline specifies what is found in the real installation,

indicating not only the model and software of each device but also its serial number and where is it located in the installation. It can also specify if a device is not installed but is in stock or in the maintenance workshop. This baseline is critical for the maintenance process, as it helps managers keep track of all installed items and make estimations of availability and maintainability.

These baselines represent the minimum requirement to manage and control the configuration of the signalling systems, but depending on the desired level of detail, more baselines can be added, such as test baselines to gather all the tests performed on an item, or delivery history baselines to trace an item's timeline (how long it has been installed and where; how long it has been in the workshop or in stock).

The database links the information from each device (both software and hardware) with the installation architecture (installed baseline), all the documents related to it (documentation baseline) and the information related to it, such as the tools needed to perform any maintenance action (system design baseline). Each item is also linked to the rest of the items in that system (system design baseline and installed baseline). The proposed baselines and their relations are shown in Figure 4.

For good configuration management, it is critical to identify the configuration items (i.e., all the information that we want to gather and control). Each item (i.e., piece of information) is placed on the related baseline and is assigned the properties that will allow us to determine the relations between it and the rest of the items (information in the database).

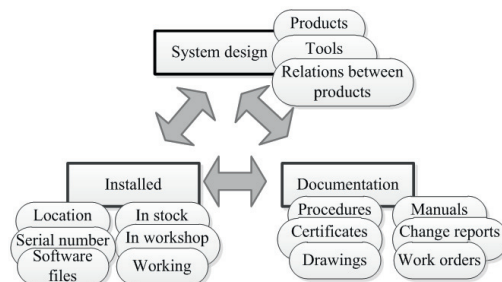


Figure 4. System configuration baselines.

In order to determine which system characteristics should be controlled and uploaded into the various baselines of configuration, such as models of the devices, tools, software documentation, etc., the different needs of the personnel involved on the operation and maintenance of the signalling system shall be taken into account. These are mainly managers from different departments (e.g. project, maintenance, safety, quality, etc.) but also the maintenance personnel. On the other hand, some documentation may not require uploading onto the database, such as internal notes or certain proceedings that are not relevant.

In the case of the different devices that are part of the signalling system, there are varying levels of detail. It is important to decide the minimum level of detail that the installed configuration baseline must achieve and which items will not be covered. Items like generic cables or fasteners do not have serial numbers,

making it difficult to control where they are placed. They are also easy to replace (not repairable).

Modern signalling systems are based on modularity and line-replaceable units (LRU) to make easier maintenance operations and increase availability. The modules can be replaced quickly in case of failure to restore the system to service while the failed module is brought to the maintenance facility to be repaired. This procedure increases availability by reducing downtime. The method proposed here takes the LRUs as the minimum level to control in the configuration baselines.

For good maintenance it is necessary to identify the configuration where a repair is necessary and to minimize the time or steps required to perform the repair. Tools include: maintenance tools, which can be software based, hardware based or a combination of both; manuals of maintenance, installation, and commissioning tests; mechanical drawings; electrical schemas; etc.

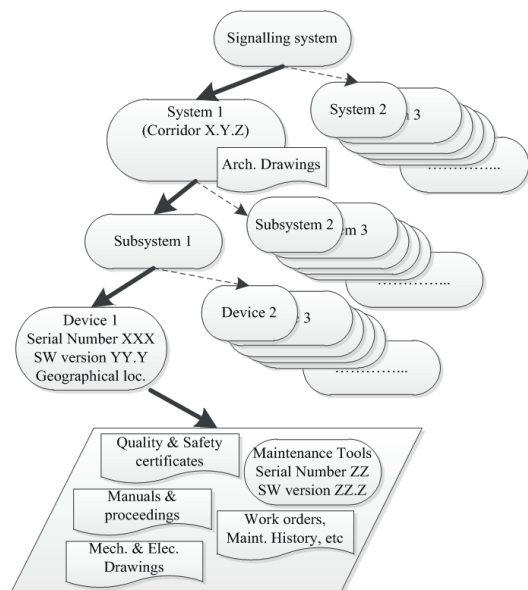


Figure 5. System configuration architecture.

To save resources and to combine the information gathered, the best option is a single database for all the company's needs. It must include such information as quality and safety certificates, traffic control, work orders, corrective and preventive maintenance, installation, configuration procedures, etc. An overview of the structure of the system configuration is shown in Figure 5.

Once the items to control in the different baselines are identified, a configuration management system which will include the storage media, the procedures to introduce, change and access the data and the tools for accessing to the configuration system must be introduced. The configurations will be managed on the cloud, based on a system of XML files interconnecting the baselines to create a unified database. There are many ways to manage the configuration system; depending on the requirements of the

specific configuration process, one solution will be better than the others.

Once a baseline is created, it should be released to all relevant personnel to inform them of any changes to the latest baseline. Before any baseline is released, it should be reviewed to avoid mistakes. The publication and distribution of the baselines make visible the real status of the system and any changes to it.

5.2 Change control management

A change can be the consequence of the deviation between the function required and the one delivered. Depending on the cause of the change, the actions needed will be managed and processed in different ways. In the case where the failure of a item is identified, it may be needed to replace the item for a new one. In this case, the change is managed as a work order and the change made is the repair or replacement of the failed device. A change can also be the result of a change in the design, such as an update in software or new equipment. The procedure defined here allows considering both repair and a design change, since the consequence of both is a change in the system. Depending on the reason, the change will be classified as maintenance or design change in the datasheet.

The change control management process includes the steps from the identification of the failure to the work order and the closure of the request. Figure 6 shows a diagram of the process.

Any failure identification should generate a change request. When a failure is identified by any person involved in the operation or maintenance of the railway line where the signalling system is located, the maintenance manager is notified of the failure.

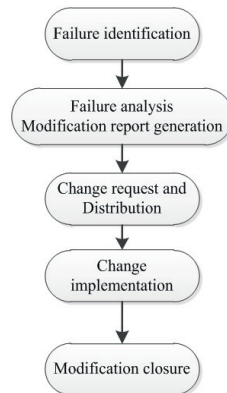


Figure 6. Change control management process.

When a change request is generated, the failure is analysed as along with the possible solutions; the modification is checked by the maintenance manager who consolidates a definitive version of the report; the change request is distributed to the personnel affected by the change, always including the person responsible for the quality and RAMS departments. The safety manager is responsible for pointing out if the change affects the safety; in this case, the change must be approved by the safety department.

The change request is reviewed by all personnel affected by the change; effects on a safety level are studied; the change request is

registered on the change control datasheet. Corrective maintenance actions and design changes are identified differently.

The change request should annex any evidence required to prove that the requested change was made and should check for non-regression (i.e., the change does not affect the functionality and fulfils all requirements for operation).

The configuration manager should be informed of the implementation status of the change in order to perform the follow up. The maintenance manager is responsible for closing the change request when all tasks have been performed and all evidence has been collected. Any change in the system should be reflected in the configuration database and linked to the change request, work orders and evidence of the change performed.

5.3 Establish integrity

The configuration should be reviewed before a new version of a baseline is created. The elements controlled in the configuration database should be verified and all modifications and changes registered in the change datasheet should be implemented in the baselines. Integrity and consistency between the configuration database and reality should be assured [10].

A review of the configuration database has the goal of verifying that an element or group of elements of a configuration which constitutes a baseline fulfils the system requirements and is consistent with the real configuration of the signalling system installed in the railway network.

These reviews involve all personnel affected by the change requests involved in changes to the baseline, as well the maintenance, configuration, quality and safety managers.

A design baseline configuration review ensures that the design of the system is the same as that implemented in the railway network, checking, e.g., the hardware and software of the different devices. An installed baseline configuration review determines whether the devices (serial numbers) included in the configuration are installed in the location where the configuration database indicates. A documentation baseline review checks that all documents regarding the configuration are linked to each item of the configuration and that they are updated to include the most recent version.

The final goal of each configuration review is to check that the configuration database reflects the reality of the installation and is consistent with the changes performed during the operation of the signalling systems.

After each review, the configuration manager publishes a report on the state of the configuration database and the changes performed. This is important as it renders visible the real status of the system and any changes that have been made.

5.4 Benefits and drawbacks

Implementing the process of configuration management following the premises of the CMMI model has both benefits and drawbacks. In order to provide a better visibility of the benefits and drawbacks they are described in the following subsections. Table 1 provides a better visibility of the benefits and drawbacks described in this section, being listed and summarized.

5.4.1 Benefits

Benefits include the fact that CM provides a common vision of the different products and processes, making it easier to identify common factors that affect the availability or maintainability of the system and decreasing the time needed to gather all the information. CM offers the possibility of storing data from different sources and locations in a unified database, making it possible to access the data no matter where they are located. The configuration management process based on the CMMI model can manage the database in a structured way.

In addition, the cost of staff in charge of the infrastructure is less due to the unification of the information in one database. Costs are reduced; because the cloud is used to manage information, equipment and a location to store it are no longer required, and maintenance resources are included in the costs of the cloud.

CM also provides better control of supplies (installed equipment is linked to devices under repair and available stock, giving more visibility quicker). For example, out of 50 devices, if 40 are installed and 8 are in stock, there must be 2 devices being repaired; no equipment can be in two states.

From a management point of view, changes are more visible, and it is easy to trace where, when and why a change has been made. This allows the simulation of possible modifications to improve the availability and maintainability of the railway line and reduce costs before any real change is implemented.

Having all the information of the systems and the maintenance equipment linked to the different failure identifications and work orders provides benefits like increased productivity (less time dedicated to identifying, finding and gathering the equipment needed), increased quality (processes are standardized and do not depend on the experience of the workers), reduced time spent on maintenance (those performing the work order will have all necessary information). All these benefits lead to improved performance.

Another benefit is the return on investment due to the reduction of costs dedicated to managing the different information databases, gathering data for each task and controlling the supplies. Better control of the system and processes gives better quality and decreases failure rate (better identification, diagnosis and prognosis of failures). The result is increased capacity and better customer satisfaction.

5.4.2 Drawbacks

Some drawbacks have to be taken into account when implementing the CMMI model for configuration management process.

A possible drawback is motivating personnel to make the extra effort required to adopt new management processes. The personnel who will gather and/or use the information should be involved in the early stages of implementation. Success or failure will depend on them, so it is important to take their needs and suggestions into account and not make them feel excluded.

While the CMMI model establishes guidelines to be followed, it does not specify an exact procedure. That implies that the model will be interpreted differently depending on the person developing it.

Table 1. Pros and Cons from the new CM process

PROS	
Unified information database	The new CM process provides a common vision of the different products and processes.
	The CM process based on the CMMI model can manage the database in a structured way
	Stores data from different sources and locations in a unified database
	The database can be placed on a server and be accessed remotely.
Ensure traceability of changes	The processes are standardized and do not depend on the experience of the workers.
	Traceability is ensured: changes are more visible, and it is easy to trace where, when and why a change has been made.
Cost reduction	CM also provides better control of supplies and change control management.
	Better control of the system and processes gives better quality and decreases failure rate (better identification, diagnosis and prognosis of failures).
Increase productivity	Decreasing time needed to gather all the information.
	Return on investment due to the reduction of costs dedicated to managing the different information databases.
Increase productivity	Increased productivity (less time dedicated to identifying, finding and gathering the equipment needed),
	Reduced downtime due to repair (those performing the work order will have all necessary information).
	The result is an improvement on the maintenance performance, increased capacity and better customer satisfaction.
CONS	
Personnel has to be involved	The personnel need to be motivated in order to get them involved on the development and implementation of the CM process.
	To implement the model will require an extra effort until the process is fully implemented.
	The improvements will not be visible until the process is fully implemented.
Interpretative model	CMMI model establishes guidelines to be followed; it does not specify an exact procedure.
	The model will be interpreted differently depending on the person developing it.
	The final users should be involved in the interpretation process so that all the needs and requirements are covered
Compromise solution	The process can become extremely complicated, and the effort required can be disproportionate to the benefits expected.
	The main goal is not to create something new, but to improve the processes of the company in order to improve the performance.
Easier failure identification	All the information of the systems and the maintenance equipment is linked to the different failure identifications and work orders.

To avoid mistakes in interpretation, it may be advisable to have a team of people to discuss the various practices involved in the management process: which items should be taken into account in the configuration and those to be excluded (for example, bolts, nuts, standard cable may not be controlled since they do not have a serial number, are not specific of any device, and are difficult to locate); the management system; the structure of the configuration; how to release the different baselines and the change reports; who should perform the reviews and audits and how this should be done.

As noted above, it is important to involve the people who are or will be affected, to ensure the best interpretation of the model. The final users should be involved in the interpretation process so that all the needs and requirements are covered.

Finally, depending on the accuracy of the interpretation of the CMMI requirements, the process can become extremely complicated, and the effort required can be disproportionate to the benefits expected. To avoid this and achieve maximum benefits, it may be useful to adapt the requirements of the CMMI model to the needs of the company where the process is applied (in the case of the signalling systems, this involves the infrastructure manager). In the particular case of the configuration management process, it is also convenient to adapt the process to the characteristics of the system; e.g., signalling systems combine SW and HW and a variety of systems in the infrastructure. It is essential to remember that the goal is not to create something new, but to improve the processes of the company in order to improve the performance.

6. CONCLUSIONS

Signalling systems in railway can be defined as a system of systems (SoS). They are composed of different systems that have their own purpose but the main functionality is given by the interoperability between them. Signalling systems are challenging to model, given the amount of information derived from both software and hardware in the various locations of the systems' many devices. Since signalling systems ensure the safe operation of the railway network, their reliability and maintainability directly affects the capacity and availability of the railway network, in terms of both infrastructure and trains. The purpose of the paper is to improve the configuration management (CM) process, providing a tool that makes the accessibility of the information and its control easier to all the people involved on the signalling systems maintenance and operation. In other words, to implement CM as a useful tool for the people involved on the operation and maintenance of the signalling system.

Not to have the proper data can lead on an increase time on corrective maintenance actions due to a wrong failure identification, which brings a lower availability of the system. Even when the failure is well identified, if the item is not repaired assuring the interoperability with the rest of the systems, the system will not be operable. Hence, a configuration management process is a must.

Signalling systems provide a challenge to the CMMI model, due to the importance of tracking and managing the information derived from both software and hardware and the varied locations of the devices. In addition, it is called upon to manage a group of systems of systems. Other approaches for CM are more focused in process improvement or on software management so they do not

fulfil the requirement of manage the configuration of both hardware and software.

As CMMI is an interpretive model focusing on software management, railway signalling systems challenge its efficacy, given the importance of the railway track and the necessity of managing a combination of information derived from both software and hardware. This paper proposes a solution based on adapting CMMI model to meet the requirements of signalling systems. The recommendations in order to improve the CM include the identification of the configuration items, along with other documentation and information relevant for the maintenance and management of signalling systems. Various baselines are proposed; the relationships between the items and the baselines are identified to link the information. Change control management processes and reviews are proposed to ensure the integrity and consistency of the configuration baselines with the real world.

By performing a good CM on signalling systems the maintenance management will be improved. A unified information database for the different stakeholders provide better visibility of the system for all of them, together with an easier traceability of any change performed. Failure identification is faster since there is a better knowledge of the system as a whole and there is a direct access to the material and information needed to perform any maintenance action (manuals, tools needed, software, etc.). For these reasons, the time consumed in the corrective maintenance is reduced. On the other hand, there is a challenge on performing a useful CM process, and to achieve that the personnel involving the different areas related to signalling systems (RAMS, quality, management, engineering, etc.) have to be involved when developing the process. The model can have different possible interpretations so surveys need to be made to ensure that the best CM process is performed, and some modifications in the process can be needed in order to adapt to the real requirements of signalling systems.

7. REFERENCES

- [1] E. Anders, "Railway Signalling & Interlocking: International Compendium". 2009.
- [2] S. Bellomo, J. D. Smith II, "Attributes of effective configuration management for systems of systems," 2008 2nd Annual IEEE Systems Conference, pp. 1-184, 2008.
- [3] J.Boardman, System of systems - the meaning of of. Proceedings 2006 IEEE/SMC International Conference on System of Systems Engineering. 2006; 2006:118-23.
- [4] H. C. Briggs, "CMMI guidelines for design engineering," A Collection of Technical Papers - AIAA/ASME/ASCE/AHS Structures, Structural Dynamics, and Materials Conference, 2008.
- [5] L. Brownsword, D. Fisher, E. Morris, J. Smith, and P. Kirwan "System-of-Systems Navigator: An Approach for Managing System-of-Systems Interoperability" (CMU/SEI-2006-TN-019). Software Engineering Institute, 2006.
- [6] CENELEC: EN 50126 Railways Applications – The specification and demonstration of Reliability, Availability, Maintainability and Safety (RAMS)
- [7] CENELEC: EN 50128 Railways Applications: Communications, signalling and processing systems - Software for railway control and protection systems

- [8] CENELEC: EN 50129 Railways Applications – Communication, signalling and processing systems. Safety related electronic systems for signalling.
- [9] CENELEC: EN 50159-2 Railways Applications: Signalling and Communications - Safety-Related Communication in Open Transmission Systems
- [10] CMMI-SVC, “CMMI for Services”, ver. 1.3, November 2010, Software Engineering Institute
- [11] CMU/SEI-2005-TN-005, Relationships Between CMMI and Six Sigma, December 2005, Software Engineering Institute
- [12] CMU/SEI-2008-TN-003, CMMI or Agile: Why Not Embrace Both!, November 2008, Software Engineering Institute
- [13] R. Dayan, S. Evans, KM your way to CMMI, *Journal of knowledge management*, 2006;10(1):69-80.
- [14] S. C. B. De, de Souza, S. R. Tilley and R. M. Newman, "A study of the documentation essential to software maintenance," *Proceedings of the 23rd Annual International Conference on Design of Communication Documenting Designing for Pervasive Information - SIGDOC '05*, pp. 68-75, 2005.
- [15] S. Dhabbi, A. Abbas-Turki, S. Hayat and A. El Moudni, "Study of the high-speed trains positioning system: European signaling system ERTMS/ETCS," in *2011 4th International Conference on Logistics, LOGISTIQUA'2011*, 2011, pp. 468-473.
- [16] J.A. Fonseca, R. Winther, B.A. Gran, G. Dahll, CMMI RAMS extension based on CENELEC railway standard. *Computer Safety, Reliability, and Security*. 2005;3688:1-12.
- [17] A. Gorod, B. Sausser, J. Boardman System-of-systems engineering management: A review of modern history and a path forward. *IEEE Systems Journal*. 2008;2(4):484-99.
- [18] IEV, “International Electrotechnical Vocabulary (IEV)”, Available at: <http://www.electropedia.org/>, 2007-04 (accessed: October 2012).
- [19] J Iglesias, “New steps in ERTMS deployment in Spain: L2 in commercial operation and ETCS in commuter lines”, *ERTMS UIC Conference*, Stockholm, 2012.
- [20] AS. Jansson, *Software Maintenance and Process Improvement by CMMI*, Master Thesis, Uppsala University.
- [21] E. Joung, K. Shin, Suggestions of development methodology for railway software, *The 23rd International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC 2008)*.
- [22] Kajko-Mattsson, M., Karim, R., & Mirjamdotter, A. (2011). Essential Components of eMaintenance. *International Journal of Performability Engineering*, 7(6), November 2011, pp. 505-521.
- [23] Karim, R. (2008). *A Service-Oriented Approach to eMaintenance of Complex Technical Systems*. Doctoral thesis, Luleå University of Technology, Luleå, Sweden.
- [24] Karim, R., Candell, O. and Söderholm, P. (2009). Development of ICT-based maintenance support services. In *Journal of Quality in Maintenance Engineering*, vol. 15, nr. 2, pp. 127 – 150
- [25] TP. Kelly, JA. McDermid, A systematic approach to safety case maintenance. *Reliab Eng Syst Saf*. 2001 3;71(3):271-84.
- [26] F. Kitahara et al, Distributed management for software maintenance in a wide-area railway system. *Proceedings of the International Symposium on Autonomous Decentralized Systems*. 1997:311-8.
- [27] L.L. Liu, Software Maintenance and CMMI for Development: A Practitioner's Point of View, *Journal of Software Engineering Studies*, Vol. 1, No. 2, 68-77, December 2006
- [28] M. Matsumoto and N. Kitamura, "Autonomous decentralized train control technology," in *Proceedings - 2009 International Symposium on Autonomous Decentralized Systems, ISADS 2009*, 2009, pp. 177-181.
- [29] Parida, A. (2006). Maintenance performance measurement system: Application of ICT and e-Maintenance Concepts, *International journal of COMADEM*, Vol.9, No. 4, 30-34
- [30] T Pasquale, “Hazard Analysis of Complex Distributed Railway Systems”, *Proceedings of the 22nd International Symposium on Reliable Distributed Systems (SRDS'03)*, 2003
- [31] M Penicka, “Formal Approach to Railway Applications”, *Formal methods and hybrid real-time systems'*, Springer, pp. 504–520, 2007
- [32] Y. Ren, Fuzzy decision analysis of the software configuration management tools selection, 2010, *Third International Symposium on Information Science and Engineering*. 2011:295-7.
- [33] J. Sutherland, Scrum and CMMI level 5: The magic potion for code warriors. *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*. 2008:466.
- [34] R.Turner, A. Jain , Agile Meets CMMI: Culture Clash or Common Cause?, *Extreme Programming and Agile Methods — XP/Agile Universe 2002*, *Lecture Notes in Computer Science* Volume 2418, 2002, pp 153-165
- [35] UNISIG SUBSET-026: ERTMS/ETCS System Requirements Specification, v. 3.2.0., ERA.
- [36] EIRENE: GSM-R System Requirements Specification, v.15, UIC.
- [37] H. Yamamoto, S. Hiraguri, N. Terada and S. Shiomi, “Signalling systems suitable for secondary lines,” *Japanese Railway Engineering*, pp. 11-13, 2011.

A handheld Maintenance Workstation: Information fusion in the aircraft – ground systems gap

Torbjörn Fransson

Saab Aeronautics
581 88 Linköping
+46734183248

torbjorn.fransson@saabgroup.com

Olov Candell

Saab Support and Services
581 88 Linköping
+4613181739

olov.candell@saabgroup.com

ABSTRACT

As the amount of data and information used and produced in the maintenance process for modern aircraft rapidly increases, conventional paper based manual processes for offline diagnostics and maintenance are getting more and more inadequate and obsolete.

New aircraft types become increasingly complex, and requirements on quick, safe and secure access to relevant maintenance data in the operational context for all types of stakeholders, will be a key factor to minimize turnaround time and enable a lean maintenance process, requirements that are very much the same for both military and commercial aircraft.

Easy and timely access to detailed and quality assured product data and logistics support information about the usage of *individual* aircraft, their subsystems and components, is also a must for effective prognostics, reliability centred maintenance (RCM) and condition based monitoring (CBM). The continuous improvement of critical aircraft capabilities, are in utilization lifecycle stages driven by feedback of in-service data and operator/maintainer information.

Security is an even more difficult topic to address as it is a moving target, since threat scenarios evolve over time. Aircraft of today are much more 'IT-systems'-like, than pure vehicles, which make them more vulnerable for unauthorized intervention and access.

The paper discuss the concept of an Maintenance Workstation (MWS) solution used at Saab Aeronautics. The approach enables data exchange and information management for operation and maintenance in close proximity to the aircraft and in the sometimes adverse environment of the flight-line.

1. INTRODUCTION

The trend towards reduced maintenance cost continues. As it has been ongoing for long time now, the low hanging fruit is already picked. What is left to do is to make more radical improvements that affect more than a particular spot in the system-of-system, on-board and on the ground that constitutes a modern aircraft system. Improvements will require an efficient and well integrated on-board and ground based infrastructure of information and communication technology, with potential for continuous improvement.

To increase availability and performance of the maintenance functions from an operational perspective, it is necessary to do provide advanced diagnostic capability as close to the source (i.e.

the aircraft) as possible. It is also of vital importance to secure the attention of the technicians at the flight-line to 'yellow light' prognostics information. Diagnostics made with the aircraft on-line will also become more important. Combined with an advanced analytics engine ('reasoner'), on-line diagnostics will also reduce the amount of information support ('tech pubs') needed to analyse a specific failure pattern.

Integrating in-flight recording functions with diagnostics and maintenance procedures, including 'man-made' information produced by the technician, will also contribute to the optimization of the maintenance loop, as well as allow for rapid updates of diagnostics and maintenance procedures based on operational user experience.

It is in this context one would consider the idea of a highly flexible handheld computer device, providing access to information support, connectivity to other (ground based) information systems and communications, as well as providing the capability to fusion information produced in the aircraft with manual input, and thereby bridging the aircraft-ground systems gap.

2. WHAT ARE THE CHALLENGES?

Requirements on maintenance operations are largely based on the cold war scenario, where planes are distributed to a large number of temporary road bases. This distribution limited the amount of equipment that could be carried to the turnaround site. It also pushed the development of embedded test functions. As a result the Saab Gripen fighter has an advanced on board fault isolation functionality. As the requirement of self containing fault isolation capability has been more relaxed and the size of required support tools has decreased, there is now a road open for new modern solutions.

These issues presented in the introduction might seem simple, as this practice is quite common in other products. What complicate the task are the aerospace domains very strict requirements on safety and security (the later especially in in the world of military aviation), as well as unique requirements on deployed operation in remote areas during considerable time spans. Safety and airworthiness is also a central issue. The risks of unauthorised access, manipulation or intervention with the aircraft must be minimised, and attempts need to be detectable and so forth, which complicates the connectivity and restricts what user are allowed to do with an external interface.

All new versions must undergo a thorough certification and accreditation process which also drives the requirement to separate airworthiness design from other design parameters.

2.1 Technology push

New maintenance equipment is also a challenge, compared to previous aircraft projects, where support equipment was a quite straightforward affair. Less complex than the aircraft and designed to meet the requirements for the entire life cycle of the plane. The ground support equipment was mostly specified, designed and manufactured within supplier domain (e.g. in Sweden by Saab or other Swedish defence industry). An example is the external memory that was used to load software and retrieve recorded maintenance data from the Saab 39A Gripen aircraft. This computer (a ruggedized PC based on a Intel 286 processor) was manufactured locally in Sweden, by what now is Saab Corporations business unit for Support and Services. .

These kind of special-to-type solutions were quite costly. Neither have, with more modern processors, development time nor airworthiness qualification costs been reduced, and time to obsolescence becomes shorter and shorter. Hence, the customer will not any longer accept the necessity to buy new expensive equipment that will be discarded after just a couple of years, which lead to a strategy of buying COTS.

2.2 Security

Security has been a growing concern during the last years. A military fighter has so far primarily been treated like a weapon system. Today the onboard computer systems and information storages makes it more like an IT system. The on board system and the data stored on board will often be subject to several classifications. There is not only military secrecy to consider but also commercial and ITAR restrictions that does not allow just any type of communications architecture. Access to information must be controlled in all parts of the chain. In addition, no one shall be able to get the information by eavesdropping.

These challenges focus on, but is not limited to needs that:

- Aircraft need to carry it own data and information relevant for operation and maintenance
- MWS enables information exchange between the aircraft and ground based ICT systems. Exchange of data is easy when the level of security is equal on both sides of the interface. Otherwise it might be prohibited.
- Provide an access point for technical personnel providing both 'read' and 'write' capabilities at the flight line. Here the MWS will need to handle different types of information and access support systems in order to be a data transporter. Communication between MWS and aircraft must be on a shielded wire or encrypted wireless.
- Storing of information in the MWS. If secret information is stored or have been stored without an approved erase, the MWS have to be handled as secret. This leads to expensive solutions for the user.

2.3 Configuration Management

In some aspects each aircraft will be unique. Even if a whole squadron is built to the identical standard, there will be different

usage in terms of hours, landings, usage etc. To enable prognostics, control of installed equipment and the life cycle and usage of that equipment must be easy to handle.

Practically all items on board an aircraft has a designated service life and as parts are exchanged over the life span. It is therefore important to keep track of the parts that require follow up. Today this is primarily done in a gigantic off board fleet management system handling all parts on board and in storage with all service history. In case the fleet management system and the computers used by the aircraft maintenance crew are in different security zones, communication can be rather complicated.

Every time a piece of equipment is changed, the information has to be stored to enable the follow up of life consumption. This goes both for equipment and parts of the aircraft and for external stores like drop tanks, missiles and sensor pods.

Even a trivial operation like moving a drop tank from one aircraft to another will require paperwork as the drop tank has impact of the service life of aircraft and pylon.

Another big issue is to keep control of all onboard software.

2.4 Software Upgrades

As the Aircraft now more or less is an IT-System, much functionality can be added by upgrading the software. Rapid development techniques and frequent updates requires good CM and also good technical solutions for data loading. Previously only a few computers were field loadable by connecting a PC or load media to the aircraft. The rest had to be removed from the aircraft and loaded in a test rig. That was a quite time consuming process that also involved relabeling of the computer with the new software identity.

The increased number of software configuration items and computers makes bench loading unrealistic.

3. WAYS FORWARD

The MWS concept as a platform gives highly flexible and virtually limitless opportunities to implement functionality that will help the technician and other stakeholders in their daily tasks.

3.1 Improving aircraft availability by introducing computerized access point at the flight line.

There are both pre-flight and post-flight activities to be performed and much time can be saved with the ability to access the aircraft with modern tools.

Pre-flight activities, like hanging of a new store or updating a flight plan or map can be done very close to takeoff.

In case of start-up problems, the MWS can be used to visualize the result from the onboard embedded built in test functions to the mechanic and also giving the possibility to further process the information for a more detailed diagnostics.

Diagnostics support can also be useful post flight. In addition readout of maintenance data that is mandatory can be done at the flight line. This might seem a bit old fashioned when there are techniques enabling fault detection in real time and on line transmittal from aircraft to ground stations. For security reasons

this is not always feasible for a military aircraft even if encrypted as a transmission leads to a disclosure of position.

The MWS can either be a standalone system to be connected with other ground support system or a networked client within.

Impact Technologies developed a tool "Reason-Pro at Wing" that was run at a PDA platform. (See ref 1). The tool gave an easy way to diagnose an aircraft. By building the functions in .net portability was early ensured. Some of the features here have inspired further work on the MWS.

3.2 Facilitating Software Upgrades

When new software is delivered, the MWS will be loaded with all software needed for a specific aircraft. The MWS will be equipped with an ARINC 615 data loader that will load the avionics computer system with all applicable software. There are several types of load scenarios. For instance:

- Upgrade of data files, maps etc that is not a part of the certified aircraft configuration. That is mostly data files that are put on an onboard memory. This is done selectively. The mechanic will select what files to load from a list in the MWS.
- Upgrade of software in a single computer. This is routinely done when a faulty computer has been replaced by one from the store as the equipment are kept unloaded in storage. Here the mechanic will select what computer to load and the source files to be loaded will be selected based on the valid configuration list for the aircraft.
- Multiple upgrades. When major upgrades of the aircraft system shall be done, it will involve change of software in several computers. This can be a bit tricky as the order in which the computers shall be loaded and restarted is important to maintain. Otherwise the startup sequence of the different computers might be corrupted and the load process has to be restarted. The MWS will in this case execute macro functions allowing several restarts of avionics computers until all equipment has received their final software.

3.3 Support Configuration Management

A concept with easy identification and configuration management of equipment and stores will save a great amount of time for the user. In the MWS concept, parts and stores will be fitted with RFID tags. A scan function to read all tags and map the information to a master list, defining all parts will enable the MWS to maintain a complete list of all installed individuals for each aircraft.

A copy of the configuration list can then be uploaded and held in storage in the aircraft, thus enabling downloading to another MWS in case that the aircraft will be served at another base.

Even if RFID is a preferred method, there are other possibilities like Bar-code reading, reading of plain text/numbers from the labels or the most primitive, manual entry.

The configuration check will be supported by a built in function in the aircraft that identifies all equipment with serial numbers embedded in the firmware. No manual check of equipment labels shall be necessary.

3.4 Job cards

All maintenance activities have to be triggered by some form of work order. Some activities, like pre-flight inspection and post flight activities are less formal but all activities that requires extra man effort, tools or spares will need some formal justification.

The job card is a generic form or work order to be used. It can be used for pre planned/scheduled services, like "Perform 100h service" or "Change APU". It is also used for unscheduled maintenance based on fault reports or test results from the plane.

The MWS will have the ability to create, maintain and finalize the job card. During the maintenance phase, checklists are linked to the job card to ensure that every part in the workflow is performed.

3.5 Fault Isolation on the flight line

Even if the aircraft is designed with a built in test, this would require that the test functions are executed from the cockpit. This can only be done when the aircraft is fully powered up, which will require ground power or in case of operating on the flight line, APU. The APU is expensive to operate. Both regarding service life and fuel. By assessing recorded data in the mass memory much of the troubleshooting can be done on battery power. If the aircraft already is powered up, a connected MWS can be used to display failure codes or mirror pilot display to the mechanic.

An interesting function is to execute test functions from the MWS. Either by controlling embedded test functions in the aircraft or even more advanced to run test functions from the MWS.

3.6 Crew communication

An interesting idea is not only to transfer data but also use the aircraft – MWS connection for voice communication. Using voice over IP, heavyweight UHF-equipment could be phased out, and also make way for more flexible connectivity to other ground-based communications systems.

3.7 Fusion of data

Some faults will be detected both by the pilot and by the embedded Built-in-test (BIT) functions. The MWS gives a possibility to fusion data by adding pilot comments about the fault situation, behavior and consequences as well as group the observations with the BIT-detections. This information can then be matched with similar observations stored in the repository to enable the benefit of previous experience.

3.8 Coordinated functional workflows and security

The data fusion functionality and coordinated, formalized, functionally controlled and digitally authenticated procedures, enables optimization of the workflow as well as enables coordinated handling of access control and systems security. Not to mention improvements in the e-maintenance domain, see ref 2 and 3.

Personal login on the MWS and authentication between MWS and aircraft will be necessary to enable individually adapted information to each technician, as well as to assure that no

unauthorized personnel will e.g. upload or download data from the aircraft.

Examples of provided advanced Information and Decision Support services could be:

- electronic log books
- individual aircraft technical information (based on actual aircraft as-maintained configuration) that is role-adapted and situation-adapted (IETP)
- auto-generation of different reports to e.g. Maintenance Management Systems and Supply Chain systems
- Security monitoring and controls

3.9 Rapid update-cycle of software and on-board data and information

The discussed approach is also an enabling factor for a closed-loop cycle for rapid updates of avionics software and on-board data/information.

The introduction at the flight-line of a computerized access point to on-board services that is integrated with ground-based information systems also enables radically improved possibilities for quick data handling. Last minute software updates and configuration changes will be possible to do in a couple of minutes. It is however essential to design an integrated information logistic flow, and there are several processes and sub-process that has to be considered, e.g.:

From the developer or OEM to the user's organisation.

If just one release was issued and no support. This would just be one shot and no problem, but complex products will be continuously upgraded both regarding new functionality, new operational procedures for maintenance, replacement of obsolete spare parts and corrections in instruction or software. All this information has to be assembled and packed in a way that the validity is clear and the service organization will have an easy mapping to the applicable aircraft.

From the ground support system to the aircraft.

Actually this might be several steps before the information reaches the aircraft. There might be one central repository holding all data while the MWS to be used from a specific aircraft will be loaded from the central repository. In this flow software and field loadable data files will be transferred from MWS to aircraft. It might also be possible to transfer pictures, manuals and other information for display on the head down displays in the cockpit.

From aircraft to Ground Support Systems.

Using the MWS, post flight data, such as Maintenance data will be downloaded. Maintenance data will also include fault codes that can be used for troubleshooting and digital log book. What cannot be analysed in the MWS, can then be transferred to other ground support systems.

From the Service organisation to the manufacturer/OEM.

This link is what a big American aircraft manufacturer called "The holy grail", as they have been chasing this function for years. The feedback loop is the best enabler to create means for improvement. It is technically no advanced function but as the operator will disclose not only the use but also potential misuse, this can be rather sensitive.

4. Conclusions

This paper might look like a long wish list. That is exactly what it is. But by introducing the MWS, we will have an enabler that simplifies a stepwise integration of helpful and significant capabilities and features in the maintenance process. With an service oriented and modularly approach, it enables data exchange and information management for operation and maintenance in close proximity to the aircraft.

Concepts like the Saab Maintenance Workstation are potential game-changers in an traditionally conservative environment. Most probably it is not best introduced in an revolutionary way, but through an *evolutionary* approach.

It is a long process to introduce new concepts. According to US Navy (ref 4), it might take between 10 and 20 years to change a paradigm and it also has to be done in smaller steps. With the MWS we have an enabler making it possible to reduce that time.

Along the discussed avenues of potential MWS development, some functionality and features will probably never leave the drawing board, as other will be rapidly launched and prove the strengths of an technically and process wise sound concept.

5. REFERENCES

- Blanchard, B. S. (2004). *Logistics Engineering and Management*. Fourth edition, Prentice-Hall, Englewood Cliffs, New Jersey.
- [1] Byington, Kalgren, Donovan, Thompson: *Streamlined Avionics PHM Utilizing Portable Information and Reasoning* IEEE Aerospace Conference 2005
 - [2] Candell, O. and Karim, R. (2008). "e-Maintenance - information driven maintenance and support", In Proceedings of FAIM.
 - [3] Karim, R., Candell, O. and Söderholm, P. (2009). "Development of ICT-based Maintenance Support Services". *Journal of Quality in Maintenance Engineering*, 15(2), pp. 127-150.
 - [4] Millar, R.C. (2009) "Defining Requirements for Advanced PHM Technologies for Optimal Reliability Centered Maintenance". IEEE Aerospace conference, 7-14 March 2009, Big Sky, Montana, USA.

Simulation as Support for PBL Contract Design

Olle Wijk
Systecon
Rehmsgatan 20
104 32 Stockholm
+46 8 459 0757

olle.wijk@systecon.se

Patric Andersson
Systecon
Rehmsgatan 20
104 32 Stockholm
+46 8 459 0766

patric.andersson@systecon.se

ABSTRACT

When negotiating a Performance Based Logistics (PBL) contract between a customer and a potential supplier, the terms need to be set appropriately in order to give the supplier the right incentive to provide logistics functions that will enable high system performance.

A Monte Carlo simulation model of the logistics scenario will give good insight to the physics involved and what can be expected to be natural variations given a statistical confidence level; such information can be of great use when formulating PBL contract terms.

In this paper a method is proposed to set the PBL contract terms appropriately by designing a penalty function $y(x)$ where x is any measurable logistics parameter(s) that can serve the purpose of measuring the degree of contract fulfilment. The penalty function $y(x)$ is proposed to be of step-wise exponential type. In this paper, backorders B , plays the role of the logistics parameter x , but the ideas presented are easily applied to any other type of measurable parameter(s). The constants of the penalty function $y(B)$ are designed to take into account the inherent variations in backorders revealed by Monte Carlo simulations. It is furthermore shown that the design of $y(B)$ is dependent upon the time period T used for monitoring the backorders.

Keywords

Performance Based Logistics; PBL contract design; Monte Carlo; Simulation; Penalty function; Cost efficient; Backorder; OPUS10; SIMLOX

1. INTRODUCTION

Performance based logistics represents a potentially cost effective method for system sustainment [3]. From the customer perspective PBL means a shift away from buying parts to instead buying performance from the supplier. PBL can be applied at system, subsystem or major assembly level. A key element in PBL is the ability to measure the system performance in a well-defined way, either directly, e.g. availability, or indirectly by measuring given logistic parameters, e.g. backorders. Monitoring and following up logistic parameters in the supply chain can on its own be a driver for supply chain performance improvements [2].

When designing a performance based logistics (PBL) contract, simulations can provide both the customer and the supplier with a better understanding of the consequences of the contract. By analysing the scenario in advance, the best possible contract can be constructed for both parties, ensuring cost efficient use of resources.

In this paper the degree of PBL contract fulfilment is proposed to be evaluated using a penalty function $y(x)$, where x is any

logistics parameter that can be measured in a well-defined way over a time period T . In Section 4 it is shown that the time period T will influence the design of the penalty function $y(x)$. In this paper the backorder measure B is used as an example for designing the penalty function $y(B)$, but the ideas can easily be extended to encompass other logistics parameters. In fact $y(x)$ could be multidimensional, i.e. x being a vector of several types of logistics parameters.

In [1] it is shown that appropriate results collected from Monte Carlo simulations enable evaluation of alternative penalty (or reward) functions $y(x)$ suggested in a PBL contract negotiation. In this paper guidelines are provided for *how* $y(x)$ should be *designed* to meet the customer and supplier objectives in a satisfactory way for both parties. Rules for constructing $y(x)$ are described in Sections 4 and 5.

The outline of the paper is as follows. In Section 2 a fictitious logistics scenario is described that is used throughout this paper to illustrate important points. Section 3 provides an initial analysis of the scenario using spares optimization and simulation, while Section 4 studies the inherent variation of backorders in more detail. Section 5 provides guidelines for an initial design of the penalty function $y(x)$ as a function of backorders ($x = B$).

In Section 6 the proposed penalty function $y(B)$ is evaluated on a validation simulation data set. Furthermore the consequence of modifying a design parameter of $y(B)$ is analysed. Section 7 discusses how sensitivity analysis can be performed on the logistic scenario.

2. SCENARIO

This paper studies the formulation of PBL contract terms between a customer and a supplier using backorders B as a performance metric. The PBL contract value C that should cover the supplier's Life Support Cost (LSC) expenses is:

$$C = 500 \text{ MUSD.}$$

The supplier responsibility is to provide a cost efficient spares stock and repair services so that the average system availability A is

$$A \geq 85 \%,$$

which can be translated to a backorder requirement, see Section 3.

The PBL contract covers a 10 year period where the average backorders are measured and monitored on a time period T basis to ensure that the supplier fulfils the contract commitments.

The support organization is shown in Figure 1 and consists of three levels, an operating base, a storage facility and a contractor workshop.

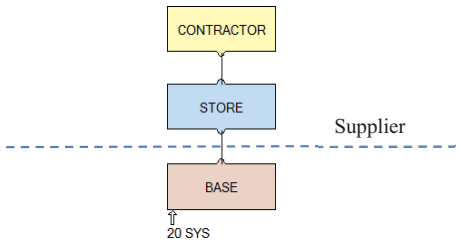


Figure 1. The Support organization.

The 20 technical systems at the BASE are utilized on average 15 % of the calendar time. Each system consists of 1000 items, out of which 400 are repairables and 600 discardables. The item repair time and lead time for reorders are 6 months. The mean time between system failures (MTBF) is 20 hours.

3. INITIAL ANALYSIS, SPARES OPTIMIZATION AND SIMULATION

An initial analysis of the scenario using the spares optimization tool OPUS10 [4] shows that the system operational requirement $A \geq 85\%$ can be met if the average number of backorders is $B = 3.14$, see Figure 2 and 3.

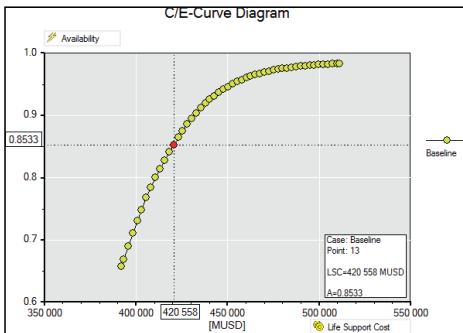


Figure 2. OPUS10 result: Cost efficient (CE) point 13.

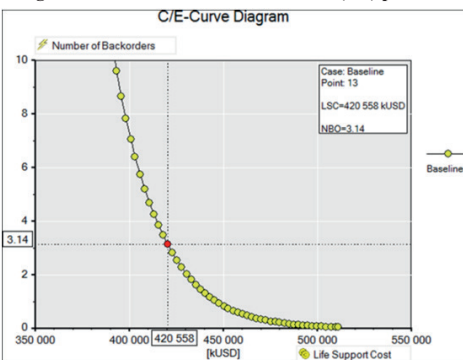


Figure 3. OPUS10 result: Corresponding backorder requirement (CE point 13), compare Figure 2.

STSIZ / Station: Stock allocation		
ID	STID: Station identifier	QTY: Total number of each station
1	LRU1	3
2	LRU2	4
3	LRU3	3
4	LRU4	4
5	LRU5	5
6	LRU6	6
7	LRU7	1
8	LRU8	7
9	LRU9	3
10	LRU10	3
11	LRU11	3
12	LRU12	4
13	LRU13	5
14	LRU14	2
15	LRU15	5
16	LRU16	4
17	LRU17	2
18	LRU18	3
19	LRU19	2
20	LRU20	3
21	LRU21	2
22	LRU22	6

Figure 4. OPUS10 result: Parts of the optimal spares stock corresponding to CE-point 13.

The optimal spares stock suggested by OPUS10 (CE-point 13) is used in the Monte Carlo based simulation tool SIMLOX [5] to verify the OPUS10 results, and provide additional information regarding the inherent variations in backorders. The simulations covers 11 years of operation out of which the backorder results for the first year are ignored to avoid the transient effects at the beginning of the simulation. The 11 year simulation is repeated 100 times using a different initial random seed in each replication. Hence the results presented in this paper are based upon backorder statistics from

$$10 \text{ years} \times 100 \text{ replications} = 1000 \text{ years of simulations}$$

In Figure 5, the backorder results from replication 1 are shown vs. time where the backorders are averaged for each 24 hour period. Figure 6 shows the same type of results from replication 12, where it is seen that the backorder variations are greater compared to replication 1. Averaging up all backorder results from the 100 replications for each 24 hour period the SIMLOX graph in Figure 7 is obtained. The total average number of backorders with respect to time and replications is $B = 3.09$ and is shown in Figure 8, this SIMLOX result is consistent with the result obtained from the OPUS10 analysis (Figure 3).

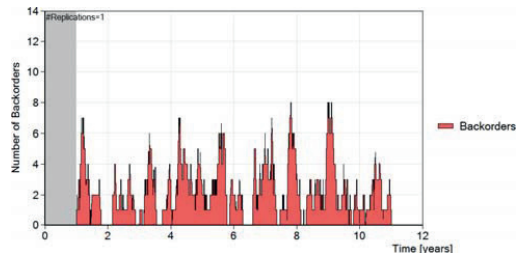


Figure 5. SIMLOX result, replication 1: backorders vs. time ($\mu = 2.2, \sigma = 1.8$).

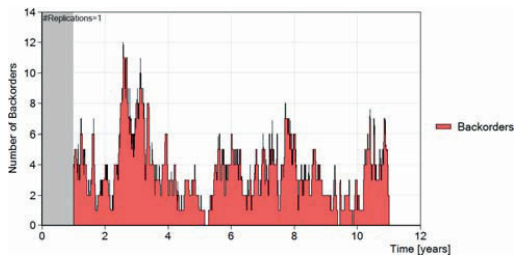


Figure 6. SIMLOX result, replication 12: backorders vs. time ($\mu = 3.6, \sigma = 2.1$).

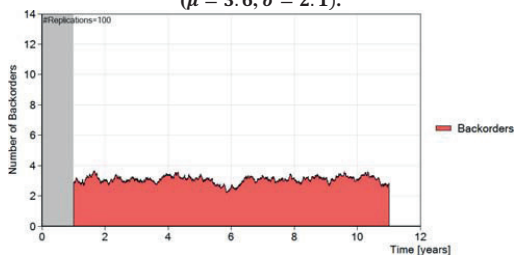


Figure 7. SIMLOX result: Backorders averaged over 100 replications.

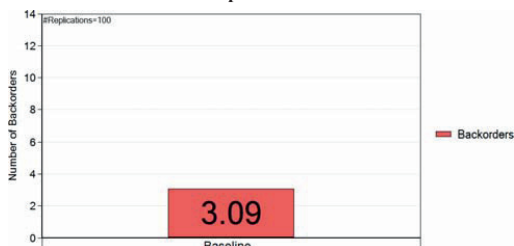


Figure 8. SIMLOX results: Backorders averaged over 100 replications and the simulation period (10 years).

4. BACKORDER VARIATIONS

From the previous section the SIMLOX Monte Carlo simulations indicate that the inherent backorders variations can be great over time. It is therefore important to consider this fact when designing the penalty function $y(B)$, see Section 5, but first the backorder variation dependence upon the measurement time period T should be considered.

The graphs in Figure 9 show the backorder probabilities $P(k \leq B < k + 1)$ evaluated for time periods *days*, *weeks*, *months*, *quarters* and *years*. The graphs indicate that the standard deviation σ of the backorders B decreases as the time period T is increased. For a daily time period T the backorder standard deviation is $\sigma = 2.1$ but for a yearly time period it is only $\sigma = 1.1$. The penalty function $y(B)$, which should take into account σ in the design, will therefore look different depending upon which time period T being used.

Once a suitable measurement time period T has been selected the design of the penalty function $y(B)$ can start. The time period T should be long enough so that the supplier has time to remedy defects in the support concept before the next measurement falls

out. However, the time period T should not be too long since then the feedback loop to the supplier concerning defects in the support concept becomes too long. In this paper a monthly time period T is selected when designing $y(B)$.

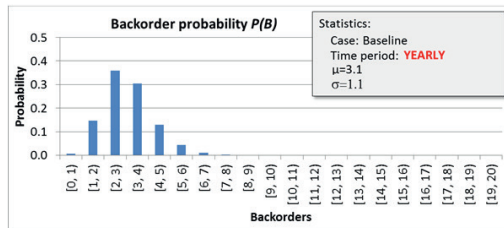
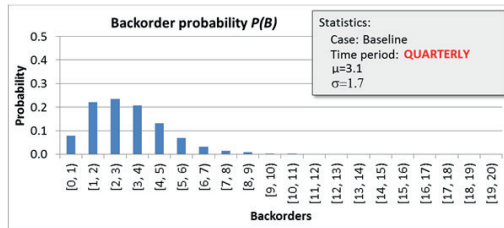
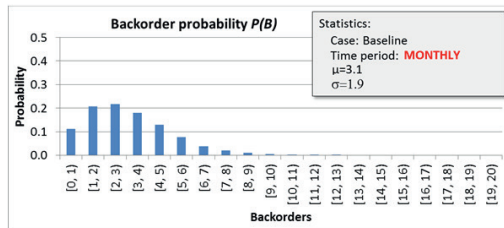
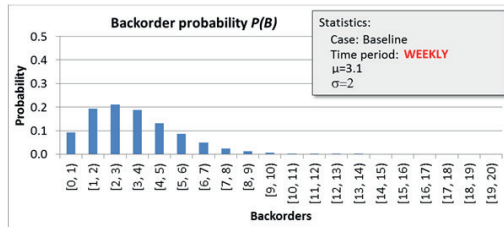
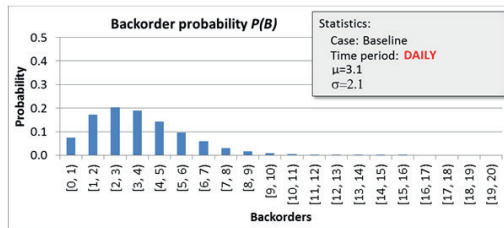


Figure 9. The backorder probability dependence upon the time period T .

5. PENALTY FUNCTION DESIGN

The penalty function $y(B)$ should not punish the inherent variations in backorders too much. By estimating the backorder probability $P(B)$ given a measurement time period T using Monte Carlo simulations as described in Section 4, the inherent variation in backorders is revealed. Two statistical measures of interest that can be obtained from the $P(B)$ estimate are the mean value μ and the standard deviation σ . An upper backorder threshold covering the most common backorder variations can then be written as

$$B_y = \mu + \alpha\sigma, \quad \alpha \in [1, 2],$$

where α needs to be selected so that the probability $P(B \geq B_y)$ is small.

Using B_y as a threshold for the penalty function $y(B)$ the following step-wise exponential function is proposed

$$y(B) = \begin{cases} \min(y_{max}, y_{min}(1 + f_y)^{\lfloor \frac{B-B_y}{\Delta B} \rfloor}), & B \geq B_y \\ 0, & B < B_y. \end{cases}$$

The penalty function $y(B)$ contains the following constants:

- y_{min} : Minimum penalty per time period T
- y_{max} : Maximum penalty per time period T
- f_y : Penalty increase fraction
- B_y : Backorder penalty threshold
- ΔB : Backorder step size

The function $y(B)$ is recommended to be designed iteratively by evaluating it on simulation results. This Section provides some rule of thumb guidelines for an initial design of $y(B)$

The constant y_{max} in the penalty function $y(B)$ represents the maximum penalty for a backorder measurement time period T . If the PBL contract covers N time periods, the total maximum penalty becomes Ny_{max} , where y_{max} should be selected so that Ny_{max} becomes a significant fraction β of the total PBL contract value C , i.e.

$$y_{max} = \beta \frac{C}{N} \quad \beta > 0$$

In this paper $\beta = 1$ is chosen. Note that the total cost for the supplier can then overshoot the total contract value C since

$$y_{max}N + LSC = C + LSC > C,$$

imposing a loss on the supplier.

Considering the baseline scenario described in Section 2 with a monthly backorder measurement period T , data from the third graph of Figure 9 gives that

$$B_y = \mu + \alpha\sigma = 3.1 + 1.9\alpha,$$

with $\alpha \in [1, 2]$, B_y is within the range

$$B_y \in [5, 7].$$

Selecting the mid-range value $B_y = 6$, we obtain

$$P(B \geq B_y) = 0.08,$$

which is a low probability indicating that the backorder threshold $B_y = 6$ is a candidate to be used in the penalty function $y(B)$, i.e. no penalty below this threshold and a step-wise exponential penalty increase if breaking above it.

Selecting the remaining penalty function constants as

$$\begin{aligned} y_{min} &= 0.05c_{max} \\ f_y &= 1.0 \\ \Delta B &= 0.5, \end{aligned}$$

causes the penalty function $y(B)$ to rise from 0 % to 100 % of y_{max} over approximately one standard backorder deviation σ (see Figure 10). The minimal penalty y_{min} starts to fall out at the backorder threshold $B_y = 6$.

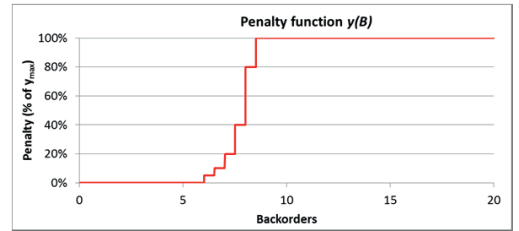


Figure 10. Penalty function $y(B)$.

6. VALIDATING THE PENALTY FUNCTION $y(B)$ USING SIMULATIONS

The penalty function $y(B)$ described in Section 5 is evaluated on backorder results obtained from a validation set of Monte Carlo simulations performed on the baseline scenario described in Section 2. Backorders are measured and averaged over a monthly time period T . Figure 11 shows the probabilities $P(y)$ for different penalty levels. Since the penalty function $y(B)$ is designed to only punish backorders B greater the inherent variations ($B > B_y$) we expect the average penalty to be low. From Figure 11 we conclude that the average penalty \bar{y} is

$$\bar{y} = \mu = 0.028y_{max},$$

with over 90 % of the time periods evaluated to a zero penalty. The total penalty \bar{Y} over N time periods is

$$\bar{Y} = \bar{y}N = 0.028C.$$

In some rare occasions, a full penalty y_{max} falls out (1.2 % of the cases) which should not trigger a big change in the supplier concept. In general, one could wait for the confirmation of at least two consecutive monthly max penalty periods before investigating the hypothesis that there could be something wrong with the support concept.

From Figure 3 it is seen that the baseline Life Support Cost is $LSC = 421$ MUSD. Since the PBL contract value is $C = 500$ MUSD we can write LSC as

$$LSC = 0.84C.$$

The total supplier cost including penalties over the N time periods covered by the PBL contract then becomes

$$\bar{Y} + LSC = 0.87C,$$

i.e. a 13 % profit margin for the supplier.

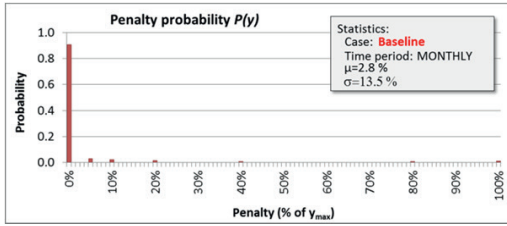


Figure 11. Penalty probability $P(y)$ for the baseline scenario.

If the supplier side of a PBL contract does not meet the operational requirements, it can be expected that the number of backorders B more often falls outside the expected backorder range. To illustrate this fact another Monte Carlo simulation is done on an *understocked* scenario using a spares assortment corresponding to 5.75 backorders on average (Figure 12). The backorder variation for the understocked scenario is shown in Figure 13, where the probability to exceed $B_y = 6$ is more significant

$$P(B \geq B_y) = 0.43$$

compared to the baseline scenario.

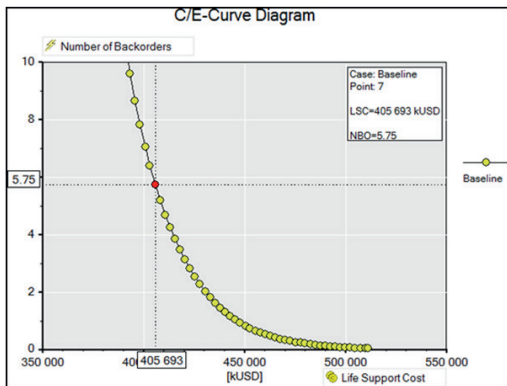


Figure 12. Selecting CE-point 7, $B = 5.75$, an understocked scenario not meeting the operational requirements.

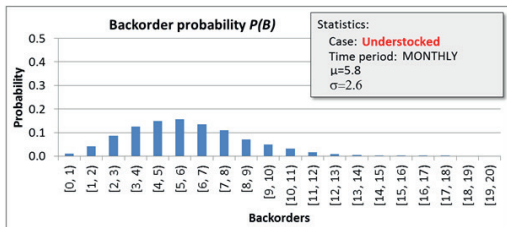


Figure 13. Backorder probability $P(B)$ for the understocked scenario of Figure 12.

Figure 14 shows the penalty probability for the understocked scenario when using the penalty function $y(B)$ described in Section 5. The average penalty \bar{y} is

$$\bar{y} = \mu = 0.22y_{max}$$

with almost 15 % of the time periods evaluating to a maximum penalty y_{max} . The total penalty \bar{Y} over N time periods is

$$\bar{Y} = \bar{y}N = 0.22C.$$

The LSC cost in the understocked scenario is 406 MUSD (Figure 12), or in terms of the PBL contract value C

$$LSC = 0.81C.$$

The total supplier cost including penalties over the N time periods covered by the PBL contract then becomes

$$\bar{Y} + LSC = 1.03C,$$

i.e. a 3 % supplier loss.

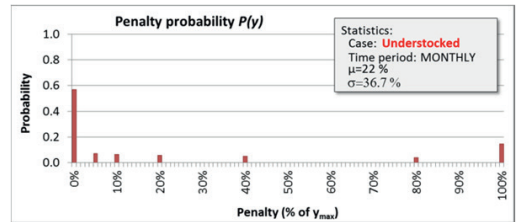


Figure 14. Penalty probability $P(y)$ for the understocked scenario.

To get a more complete overview of the consequences using the penalty function $y(B)$ the analysis done in the baseline and understocked scenario is repeated for all the cost effective assortments of Figure 12. The result of this analysis is seen in Figure 15 which also displays the upper and lower 90 % percentiles. The minimum is obtained at CE-point 15 close to the baseline stock (CE-point 13). For understocked scenarios the supplier cost $\bar{Y} + LSC$ increases rapidly to well exceed the contract value C imposing a net loss on the supplier.

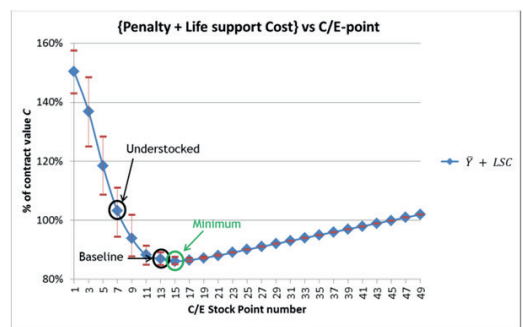


Figure 15. Supplier cost evaluated for the cost effective points of Figure 12.

Note that once all simulations of interest are done, no new simulations are needed to evaluate other candidate types of penalty functions $y(B)$. In the graph of Figure 16 the supplier cost

is evaluated using a backorder penalty threshold $B_y = 4, 5, 6, 7$ and 8, where $B_y = 6$ corresponds to the baseline scenario. Overstocking is encouraged in the case of $B_y = 4$ because otherwise the supplier risks losing money although meeting the system availability requirements. Understocking is encouraged in the case of $B_y = 4$ because the supplier can in fact increase the profit by understocking a bit (minimum cost is obtained at point 11, compared to point 13 for baseline scenario). An appropriate value of B_y should be somewhere in the range $B_y \in [5,7]$.

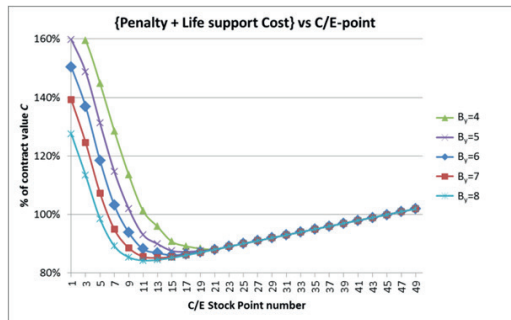


Figure 16. Supplier cost when using backorder penalty thresholds $B_y = 4, 5, 6, 7$ and 8.

7. ASSESSING RISKS

Assessing the risks before agreeing to PBL contract terms is important. From the supplier perspective it can be of interest to do a sensitivity analysis with respect parameters that involve some degree of uncertainty. To exemplify, the supplier wants to analyse what an increase in repair times would mean in lost profit.

In Figure 17 the result of simulating the a scenario with 20 % increase in repair times but still using the stock suggested for the baseline scenario (CE-point 13 of Figure 3). The 13 % profit for the baseline scenario has now instead turned into a 2 % loss measured in terms of the contract value C .

The supplier can plan for various actions to take in this situation, e.g. try to reduce repair times or increase the spares levels. A spares optimization tool can be of great benefit in this decision process.

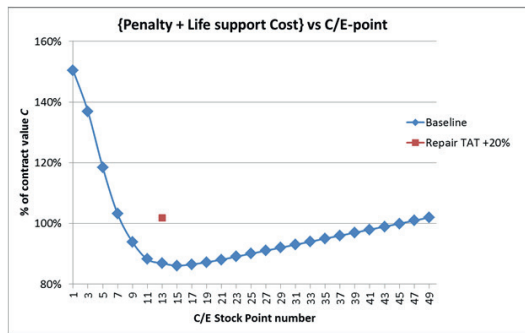


Figure 17. The impact of a 20 % repair time increase.

8. SUMMARY

In this paper key principles are discussed when designing a penalty function $y(x)$ to be used in a PBL contract agreement. Monte Carlo simulation is an important tool in the design process of $y(x)$ since it can provide statistics of the inherent variations of the logistic parameters x . The penalty function $y(x)$ is proposed to be of step-wise exponential type.

9. REFERENCES

- [1] Wijk, O., Olinger T., Hell R. (2011). "Simulation as support for decision making in negotiations". 14th annual System Engineering Conference in San Diego, USA. Available at: www.systecon.se/documents/Simulation_as_support.pdf
- [2] Hoek, R. I. van (1998). "Measuring the unmeasurable" - measuring and improving performance in the supply chain. Supply Chain Management, 3 (4), pp. 187-191
- [3] Mahon, D. (2007). "Performance-based logistics: Transforming Sustainment". Journal of Contract Management, 5 (1), pp.53-71
- [4] OPUS10 - Logistic Support and Spares Optimization. www.systecon.se
- [5] SIMLOX – Simulation of Logistics and Operations. www.systecon.se

Applying an Integrated Systems Security Engineering Approach within Aviation eMaintenance

Tim Tinney
Saab Support and Services
581 88 Linköping
+46734371726

Tim.Tinney@saabgroup.com

Olov Candell
Saab Support and Services
581 88 Linköping
+4613181739

Olov.candell@saabgroup.com

ABSTRACT

Whenever there are complex technical systems, there will always be the drive to incorporate and integrate new technologies, and this is true today with aviation eMaintenance solutions. By doing this, there can be, and most likely will be, security associated problems and challenges with both the new technologies and with the integration of those newer technologies with existing legacy technologies. Systems security engineering is an excellent capability to tackle these problems and challenges by providing a bridge between systems engineering and security engineering. This paper explores the potentiality of applying systems security engineering, based on ISO/IEC 21827, to the eMaintenance domain. In addition, this paper concludes with highlighting some of the more prominent security challenges associated with Commercial Off-the-Shelf (COTS) system solutions.

Keywords

eMaintenance, COTS, systems security, security engineering

1. INTRODUCTION

Today's actors within the domain of maintenance and in-service support related to modern aircraft and other complex technical systems, are facing major challenges related to management of ever-increasing information flow and systems complexity. Both military and commercial operators need to reduce aircraft downtime and maintenance man-hours. Heterogeneous information systems and manual information management have the opposite effect, inducing unnecessary costs and affect efficiency negatively.

The address of these challenges can often be facilitated by the utilisation of new and innovative Information & Communication Technology (ICT), manifested in emerging approaches such as eMaintenance. It is long known that properly applied, ICT has the potential to facilitate business information support in terms of product support, the maintenance process and alignment of the maintenance processes with other business processes [1]. Another important foundation for improvements related to information management and support, is the ability to utilise existing ICT legacy systems and information resources, both within aircraft and on the ground. It is also necessary to coordinate the acquisition, design and production of new digital information with legacy information [2].

Though many of these legacy ICT systems have been around for decades and are based on obsolete code, data architectures, and networks [3,4,5]. Security has been a dormant problem that has

existed with product oriented legacy systems for decades in the form of latent code defects.

At the same time modern ICT based service-oriented systems, including many eMaintenance solutions, are available through various network solutions, such as via an Internet Protocol (IP). The implementation of these internet-based solutions within aviation, as well as in other domains, requires appropriate measures to improve the safety and security aspects of a system. Troublesome examples of this fact are critical industrial processes like gas pipelines and fresh-water supplies, which are accessed, monitored and remotely controlled through the use of internet and other internet-connected networks. Because of this, security has become of paramount importance due to overt attacks affecting system functionalities [6].

However, to address this situation, the application of systems security engineering practises, as it is discussed in ISO/IEC 21827, needs to be integrated within the eMaintenance domain in a systems engineering (SE) perspective, as delineated in ISO/IEC 15288. The reason for this is due to SE emphasising a system-of-systems view over the whole life cycle, as well as its universality and focus on design and application capabilities from multiple engineering disciplines and levels of technology.

Hence, this paper focuses on the some benefits of using systems security engineering principles within the eMaintenance domain, and explores some of the challenges of maintaining and operating COTS systems in today's environment. Even though these challenges are more pertinent for the aviation domain, they can directly or indirectly be associated with similar challenges within the civil, environmental and natural resources domains.

2. EMAINTENANCE SECURITY ASPECTS

Growing and poorly integrated information management (IM) in heterogeneous ICT environments, often specific to aircraft type or supplier, such as computer based support systems and operator Maintenance/Fleet Management Systems, drives cost, time and resources for operators of military and civil aircraft, as well as having a negative impact on aircraft and fleet availability.

One trend within eMaintenance is the service integration of very large number of ICT devices [7] and the integration of the development of aviation information support products and services with traditional SE processes [8,9]. This includes internet enabled and cloud-based collaborative services that provide new and more efficient product support functions and service systems through integration of on-board aircraft functionalities and ground

capabilities, providing services and decision support in aerospace operation and maintenance processes [10].

Though at the same time the introduction of elaborate and integrated ICT based eMaintenance solutions raises significant security challenges in environments where systems are used in environments that imply very high requirements on safety and credibility [11,12,13], and where security has been a growing concern during the last years.

One example is that of on-board computers in modern aircraft. This has so far primarily been regarded as avionics, though in modern designs with their multiple on-board computers connected through multiple data-buses and with high capacity information storage, shows many similarities with various ground based high-tech information systems. The on-board system and the data stored inside the aircraft will also often be subject to several classifications. Hence, access to information must be assured both within the system-of-interest as well as in the enabling system, in all parts of the maintenance process (see Figure 1) and supply chain. In addition, the system integrity needs to be assured, i.e. no one should be able to access or manipulate the information at any instance.

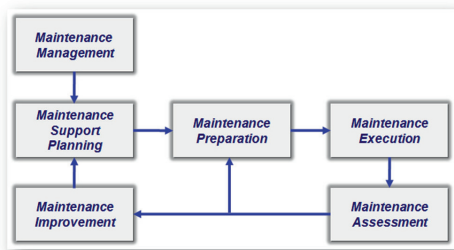


Figure 1. A generic maintenance process [14].

From an aviation operations and maintenance perspective this raises some central aspects that relate to issues of security for an eMaintenance solution as follows:

- The aircraft becomes a container and carrier of significant amounts of data and information relevant for its own operation and maintenance.
- Increased information and data requirements and traffic between the aircraft and ground based ICT systems.
- Technical personnel need to be provided, in the vicinity of the flight-line, with some type of access for interaction with data and information stored both in the aircraft and in ground based ICT systems.
- Data and information related to operation and maintenance needs to be exchanged with the aircraft, (partly) processed on-site (e.g. adding manual input) and communicated with stakeholders off-site, preferably real time (or close to).
- Components, data and information need to be secure in all aspects during the maintenance cycle through removal, packaging, handling, transportation and

maintenance/repair at operational level workshops, maintenance centres, original equipment manufacturers (OEM) or other supply chain actors.

3. AN INTEGRATED SYSTEMS SECURITY ENGINEERING APPROACH

Today the systems security terminology is primarily associated with ISO/IEC 21827, the Systems Security Engineering Capability Maturity Model (SSE-CMM). This ISO/IEC was originally prepared by the now International Systems Security Engineering Association (ISSEA) during the late 1990s as a process reference model, i.e. a roadmap for establishing and maturing security practices. The SSE-CMM can be used by a variety of organizations, for example by security service providers, countermeasure developers, or by product developers. Essentially, the SSE-CMM is a tool to evaluate security engineering practices and define improvements, a method for security engineering certifiers and evaluators to establish confidence as an input to system or product security assurance, and to provide a standard mechanism to evaluate the provisioning of security engineering capabilities. [15]

One instance where systems security concepts were skilfully used, and in accordance with ISO/IEC 21827, was by the US Department of Defense (DoD) in 2005. Here the US DoD pushed for better software and system assurance measures to combat the threat of malicious software tampering. Their focus on systems security deftly brought to light the need and benefits of systems security engineering [16]. US DoD also needed to account for vulnerabilities in system hardware, firmware, and integration. As a result, the concept of system assurance was developed as “the justified confidence that the system functions as intended and is free of exploitable vulnerabilities, either intentionally or unintentionally designed or inserted as part of the system at any time during the life cycle.” This though seems to be more of a goal than a reality in many areas of aviation today.

One of the conceptual pillars within systems security engineering is that it is composed of multifarious security sub-disciplines. In ISO 21827, it is stated that it is likely that the following specialty security sub-disciplines could be required, as an example: operations security, information security, network security, physical security, personnel, administrative security, communications security, emanation security and computer security. These are only examples. There may be a requirement to incorporate other security sub-disciplines. This concept of interweaving security sub-disciplines is extremely valuable and is required when conducting a thorough systems security analysis and security risk assessment of an aviation legacy system.

To better convey the interrelatedness, or connectivity, and potential synergies between security sub-disciplines, Figure 2 is provided. Some documentation refers to the above security sub-disciplines as branches of security or even areas of security. Either way, each security sub-discipline focuses on a specific security need. When taken from a holistic viewpoint, there are essentially a number of avenues by which a system can be targeted. Thus, to provide a coordinated effort to secure any system, multiple security sub-disciplines are required

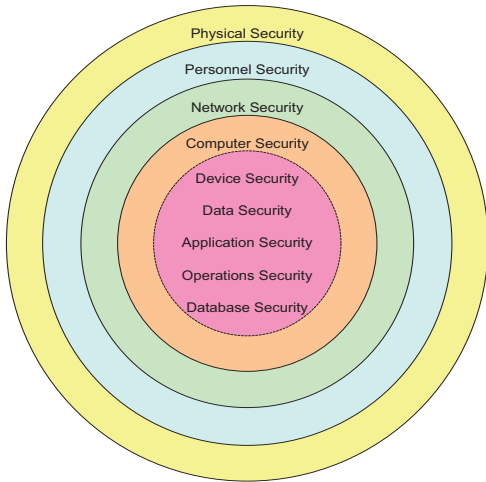


Figure 2. Abstract depiction of security sub-disciplines to show potential interrelatedness [17].

This is where systems security engineering can be of value. As a concept, it is a bridge between system engineering and security engineering. Systems security engineering uses security engineering principles and concepts to provide for and assist with organizing and understanding a system's requirements and processes as it pertains to the entire life cycle and the entire organization, as an example. It helps to coordinate a focused multi-security sub-discipline effort where needed. As delineated in ISO/IEC 21827, this is accomplished through a number of generic capability levels (per Figure 3, below) and model process areas with corresponding generic predefined security base practices (per Figure 4, below).

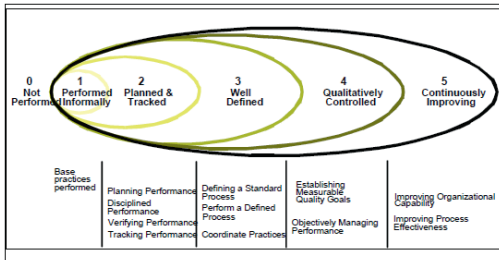


Figure 3. Generic capability levels represent the maturity of security engineering organizations [18].

The capability levels help in performing an assessment and implementing improvements of an organization's process capability. They help define an organization's maturity. By accomplishing these, an organization gains dimension of where it is relative to its business goals, as well as prioritization and sequence of its activities.

Security Engineering Process Areas	# of Base Practices	Project and Organizational Process Areas	# of Base Practices
Administer Security Controls	4	Ensure Quality	8
Assess Impact	6	Manage Configuration	5
Assess Security Risk	6	Manage Project Risk	6
Assess Threat	6	Monitor and Control Technical Effort	6
Assess Vulnerability	5	Plan Technical Effort	10
Build Assurance Argument	5	Define Organization's Security Engineering Process	4
Coordinate Security	4	Improve Organization's Security Engineering Process	4
Monitor Security Posture	7	Manage Product Line Evolution	5
Provide Security Input	6	Manage Systems Engineering Support Environment	7
Specify Security Needs	7	Provide Ongoing Skills and Knowledge	8
Verify and Validate Security	5	Coordinate with Suppliers	5

Figure 4. An overview of model process areas and corresponding number of base practices [19].

Within the confines of ISO/IEC 21827, there are 129 base practices, organized into 22 process areas. The resultant base practices cover all major areas of security engineering, as well as address project and organizational domains. They have been drawn from systems engineering, Software Capability Maturity Model, and a wide range of existing material, tested practices, and expertise.

It is at this level of integration that systems security can and should be incorporated in an organizations business practices. This will facilitate organizations to begin to actively tackle the many challenges of performing viable system's assurance analysis of their aviation legacy systems as part of their routine eMaintenance operations, instead of living in a world of denial. The key is to know what the security risks are and decide how to deal with them. Systems security engineering does this as well as provides a framework for organizations to handle their product security challenges.

4. AVIATION SECURITY CHALLENGES

One of the more convoluted long standing security challenges today is our use and dependence on Commercial Off-the-Shelf (COTS) systems in aviation. The use of COTS has allowed industry and the military to incorporate new technologies into systems more quickly than normal developmental programs and processes. Its use can reduce research and development costs, but often at the expense of security. Some of these problems have been well documented since the 1990's [20,21]. Since then, the problems keep piling up. With the advent of the Stuxnet virus in 2010, which plagued Siemens systems software, there is now concrete evidence of system vulnerabilities being exploited by way of sophisticated designed attacks. These problems make COTS maintenance over time a potential mine field of security related problems.

This begs the question, if there are long standing security challenges with COTS aviation systems, then why are they still being utilized in today's new aircraft designs and manufacturing? To answer this question, it must be understood that most, if not all, COTS aviation systems are very specialized, such as a flight management system. This specialization does not come cheaply, although if any one aircraft manufacturer were to venture to invest in such a technological development, it would definitively cost more. This has led to a very commercially competitive environment and has essentially created two main vulnerabilities.

4.1 Hidden complexities

The first vulnerability is hidden complexities. Generally COTS technologies and standards were available to the public, but now they are transitioning to more proprietary technologies using complex layers. This is occurring unabated for hardware, applications and operating systems [22]. This is not surprising, by any means. This can and has progressed out of control, though.

An example of this is an incident in August 2006 where a Malaysia Airlines jetliner was flying from Perth, Australia, to Kuala Lumpur, Malaysia. A defective software program supplied incorrect data about the aircraft's speed and acceleration. The confused flight computers sent the Boeing 777 on a 3,000 foot roller-coaster ride. Essentially, aircraft software programs have become too large (some with more than five million lines of code) and too complex to be adequately tested and are fielded without any guarantee that they will always work. [23]

What often occurs with such complex systems is that more time goes into integrating the COTS systems than planned. The result is that if there was any planned requirement for systems assurance testing and verification, it usually is either put off to a later date or forgotten completely.

In addition to this, too many times engineers and management use the reasoning that, "if it is not broken, then do not fix it." What is forgotten or overlooked are the problems with software latent errors. A good example of this was the space shuttle which has been held up as a paragon of software-development. Even though its code had been reviewed many times, there were cases of vulnerabilities and bugs from code that had been written ten or twenty years earlier. Tracking the number of vulnerabilities found in any software product might be more indicative of how aggressively the code is being investigated for vulnerabilities, rather than the base rate of vulnerabilities that exists. Old code can be found to be vulnerable to new categories of vulnerabilities as they are discovered, not to mention vulnerabilities resulting from to new technologies. [24]

4.2 Function-rich products

The second vulnerability associated with COTS software products, stems from the inevitable competitive practice of providing function-rich products that are not necessarily secure. There are also many COTS products that come with "fringe benefits" that are basically undocumented features and functions. Here is a classic case of where too many options is just as bad, if not worse, than not having enough. As a result, the development and implementation steps to incorporate these COTS products into larger systems must take concerted efforts to identify these fringe benefits and unused functionalities to ensure that they are controlled and contained. [25] This type of work costs time and money and can often raise more security related problems like raise the risk of an increased number of latent defects.

The next security challenge to be discussed is Stuxnet. This is a computer worm virus that was spread by devices plugged into USB computer ports and effected industrial control system devices. The actual effects of Stuxnet are up for conjecture at this point. There are reports that Stuxnet had the ability to degrade or destroy the software on which it operated, while other reports from Iran stated that the virus was designed to transfer data about Iranian production lines to locations outside of Iran. As of 25 September 2010, Iran had identified IP addresses of 30,000 industrial computer systems that had been infected. [26]

4.3 Stuxnet

Iran seemed to be the most effected by Stuxnet, from all reports. This even though there were reports of varying degrees of disruption from Stuxnet in Indonesia, India, Pakistan, Germany, China, and the United States [27]. In most cases, the actual damage could have been worse since most countries did not publicize the effects from Stuxnet. The main point here is that Stuxnet was a tailored computer worm virus that was specifically designed to attack Siemen's industrial control systems by using a existing system vulnerability.

This type of incident should provide the needed ammunition for governments and industry to look closer at their COTS systems in a different light, especially those systems that are responsible for the provisioning of our electricity and water. A more recent paper on these types of attacks was published by the University of Cambridge by Leverett [28]. It does not matter if a system is open or closed looped or whether it is connected to the Internet.

4.4 Market domination

Lastly, the third and final security challenge to be discussed in this paper is the basis for many of the problems associated with product related security. This is in reference to the practice of companies whose primary focus is to dominate platform markets by shipping software and products with too little security. Once a company has achieved some level of domination, then it adds security, often times the wrong kind [29]. This is actually a self-inflicted wound, i.e. from the software industry's own developmental processes.

Knowing how to circumnavigate this type of bad business practice is one thing, but the real problem is how to deal with the existing product systems and solutions that were fielded previously, i.e. our current legacy systems. This is where there is a need for a better overall understanding of an asset's existing interrelated systems security posture as it pertains to all applicable security sub-disciplines.

The aforementioned security challenges are but some of the many challenges and considerations that must be dealt with when performing eMaintenance, as they pertain to COTS solutions. There are many more challenges beyond the realm of COTS that must also be considered.

5. CONCLUSIONS

There is a need for enhanced integration, like in the approach of eMaintenance, of the system-of-interest (e.g. a fleet of aircraft), the customer's enabling systems (e.g. maintenance and support system), as well as with the ICT-system provider regarding data and information exchange. This will help with exploiting technological development of maintenance and support related data and information produced internally in a system-of-interest and its enabling systems. New business models like PBL, and continuously increasing requirements on improved dependability and operational flexibility for complex technical systems, will push the trend within the maintenance industry towards more integrated maintenance service packages. Though, these initiatives will not necessarily go hand in hand with emerging threats and new requirements on systems security engineering.

The information management in most maintenance processes interacts with a number of correlated processes in an enterprise, e.g. business, operation, information, support, and logistics. Hence, a systems security engineering approach to the eMaintenance domain should incorporate methods and tools

necessary to address the challenge of a conglomerate of actors, processes and ICT based systems.

Systems security engineering provides for and assists with organizing and understanding a system's requirements and processes as it pertains to:

-the entire life cycle;

-the entire organization;

-concurrent interactions with other disciplines: such as system, software, hardware, human factors, communications & test engineering; operations & maintenance;

-all security sub-disciplines.

One of the primary objects that all of industry and the military should strive for is to know what the security risks are for any and all applicable systems. Once the risks are known, the decision how to appropriately deal with them. Living in a world of risk denial is not only dangerous, but ethically immoral. Systems security engineering does provide some viable guidelines and base practices that can assist with providing a framework for organizations to handle their product security challenges.

Lastly, there are many security challenges that may not be commonly known to all as they pertain to the eMaintenance domain. The security related challenges previously discussed with COTS systems are but a few. With today's connected world, little problems in one area of the world can spiral out of control very quickly. Governments and industry need to work closer together to combat these challenges. What once was thought to be safe and secure has quickly proven to be vulnerability the next day. The advent of new technologies has often been regarded by management as a business enhancement. In reality new technologies quite often create more system vulnerabilities.

6. REFERENCES

- [1] Williams, T., Maull, R., Ellis, B. (2002). Demand chain management theory: constraints and development from global aerospace supply webs, *Journal of Operations Management*, vol. 20, iss. 6, pp.691-706.
- [2] McDonnell, M. and Clegg, B. (2007). Designing a Support System for aerospace maintenance supply chains. *Journal of Manufacturing Technology Management*, vol. 18, no. 2, pp. 139-152.
- [3] Seacord R., Plakosh D., and Lewis G. (2003). *Modernizing Legacy Systems: Software Technologies, Engineering Processes, and Business Practices*. New York: Addison-Wesley.
- [4] Ulrich W. (2002). *Legacy Systems: Transformation Strategies*. Upper Saddle River, NJ: Prentice Hall PTR.
- [5] Mitleton-Kelly, Eve (2006) IT legacy systems: enabling environments that reduce the legacy problem: a complexity perspective. In: Madhavji, Nazim and Fernandez-Ramil, Juan and Perry, Dewayne, (eds.) *Software evolution and feedback: theory and practice*. John Wiley & Sons, pp. 53-70.
- [6] Ijure, V. M., Laughter, S. A., Williams, R. D. (2006), Security issues in SCADA networks, *Computers & Security journal*, Vol. 25, Issue 7, pp. 498-506.
- [7] Zhang, W., Halang, W., Diedrich, C., "An agent-based platform for service integration in e-maintenance", *IEEE International Conference on Industrial Technology*, Vol. 1, pp. 426-433, 2003.
- [8] Candell, O. (2004). Development of user information products for complex technical systems. Luleå tekniska universitet. Licentiate thesis, Luleå University of Technology; No. 2004:25)
- [9] Candell, O. 2009. Development of information support solutions for complex technical systems using eMaintenance. Doctoral thesis, Luleå University of Technology.
- [10] Careless, J. (2004). Airbus's New Support Strategy Values Customer Service. *Aviation Maintenance*, April 1, 2004, available at: <<http://www.aviationtoday.com/am/categories/maintenance/170.html>>, accessed: 27 September 2009.
- [11] Wachs, M. (2002). *Fighting traffic congestion with information technology*. *Issues in Science and Technology*, vol. 19, no. 1, pp. 43-51.
- [12] Graber, M. (2004). *The Safety of Computer-Based Medication Systems*/In reply. *Archives of Internal Medicine*, vol. 164, no. 3, pp. 164, 339.
- [13] Bischoff, Hans-Jürgen. (2008) Risks in modern society. Volume 13 of Topics in safety, reliability, and quality. IFIP transactions. Computer science and technology, vol. 33, p 17. Springer.
- [14] IEC (2008). 60300 (3-16): Dependability management - Part 3-16: Application guide - Guideline for the specification of maintenance support services (Final Draft). International Electrotechnical Commission, Geneva, Switzerland.
- [15] ISO/IEC 21827:2008(E). Information technology-Security Techniques-Systems Security Engineering-Capability Maturity Model (SSE-CMM). Geneva, Switzerland: ISO.
- [16] Baldwin, K. (2009). Systems Security Engineering: A Critical Discipline of Systems Engineering. *INCOSE INSIGHT*, July 2009 issue. Vol. 12 Issue 2, pp. 11-12.
- [17] Jacobs, S. (2011). *Engineering Information Security*. Wiley & Sons, Hoboken, New Jersey.
- [18] ISO/IEC 21827:2008(E). Information technology-Security Techniques-Systems Security Engineering-Capability Maturity Model (SSE-CMM). Geneva, Switzerland: ISO.
- [19] Richardson, J., Bartol, N. and Moss, M. (2004). *ISO/IEC 21827 Systems Security Engineering Capability Maturity Model (SSE-CMM), A process Driven Framework for Assurance*. <<https://buildsecurityin.us-cert.gov/swa/downloads/3-ISOIEC-21827-CMMIAndAssuranceAug2-Moss-Richardson.pdf>>, accessed: 2 August 2012.
- [20] Alford, L. (1999). The Problem with Aviation COTS. *Acquisition Review Quarterly*. Summer 1999.
- [21] Voas, J. and Jeffery Payne. (1998). COTS Software Failures: Can Anything be Done. *Proceedings IEEE Workshop on Application Specific Software Engineering and Technology*.
- [22] Jacobs, S. (2011). *Engineering Information Security*. Wiley & Sons, Hoboken, New Jersey.
- [23] Bonabeau E. (2007). *Understanding and Managing Complexity Risk*. MIT Sloan Management Review. Summer 2007.

- [24] Shostack, A. and Stewart, A. (2008). *The New School of Information Security*. Upper Saddle River, NJ: Addison-Wesley.
- [25] Herrmann, D. (2007). *Complete Guide to Security and Privacy Metrics*. Boca Raton, FL: Auerbach Publications.
- [26] Rebane, J. (ed) (2011). *The Stuxnet Computer Worm and Industrial Control System Security*. New York: Nova Science Publishers.
- [27] Rebane, J. (ed) (2011). *The Stuxnet Computer Worm and Industrial Control System Security*. New York: Nova Science Publishers.
- [28] Leverett, E. P. (2011). *Quantitatively Assessing and Visualising Industrial System Attack Surfaces*. University of Cambridge, Computer Laboratory, Darwin College.
- [29] Anderson, R. (2008). *Security Engineering: A Guide to Building Dependable Distributed Systems*. Indianapolis, IN: Wiley Publishing, Inc. 2nd ed.

Augmented Reality as a New Level for Maintenance Efficiency

Rúben Oliveira
CEMUC
PORTUGAL
ruben.oliveira@dem.uc.pt

José Torres Farinha
CEMUC
PORTUGAL
torres.farinha@dem.uc.pt

Inácio Fonseca
CEMUC
PORTUGAL
inacio.fonseca@dem.uc.pt

Luís Andrade Ferreira
FEUP
PORTUGAL
lferreir@fe.up.pt

Fátima Armas
ISR
PORTUGAL
mfatima.armas@gmail.com

ABSTRACT

Nowadays, the Augmented Reality (AR) can be found in several applications: to display virtually products in a 3D animated way; to display a 3D version of pictures in books; to present advanced technical instructions; to support military maintenance; to aid navigation system on nuclear power plants, issuing warnings about the dangers along the paths; to support the assembly process of medical (surgical) equipment; as a medical tool for measuring human organs.

In particular, AR should be taken into account on the development of new applications in the maintenance sector. Through AR, it can be showed to the trainee, the steps to perform an intervention and giving warnings in order to avoid mistakes, and correcting and suggesting how to handle components. Simultaneously, with this type of training, it is possible to evaluate the performance of the trainee in real-time.

The benefits of AR also depends on how information is provided. An AR application should identify the components, virtually, on the field of vision of the technician and demonstrate how assembling or disassembling the part is done. An AR application, to achieve better results, must be combined with other tools, like information systems, in order to enhance the capabilities of the technician on the field to optimize the maintenance intervention.

To improve the effectiveness of AR, the parts should be identified without using markers on them.

This paper presents this subjects and some related work that is complementary to this, with some applications and other promising results.

Keywords

Augmented Reality; Computer Vision; 3D Vision; Maintenance.

1. INTRODUCTION

This paper is related to new technologies to be applied mainly on the maintenance sector, through introduction of Augmented Reality (AR) on maintenance interventions and presenting their benefits and difficulties of implementation on industrial environments.

The AR is a technology already well implemented in education, entertainment and advertisement sectors because there are fewer requirements comparatively to the industrial environments, [1]. However, many projects have been developed for maintenance and manufacturing services, despite with limited functionalities; those projects achieved promising results that validates the continuous investments to develop better AR tools and anticipating its implementation, on large scale, on industry in the future.

These new AR tools, to support maintenance interventions, will enlarge the trainees knowledge and enhance their efficiency by minimizing committed errors and intervention times. This technology can also be applied to technical training, showing the trainee, the steps to perform an intervention on a virtual environment and indicating how to handle components. At the same time, the system is able to evaluate the performance of the trainee in real-time.

A new AR tool for maintenance sector is proposed, based on previous projects, which applies the benefits reached by them and added new functionalities to correct errors and fill the gaps of these projects.

Although all efforts in R&D made by universities and industrials on AR, it wasn't developed yet a solution robust enough and user friendly to be implemented on industrial environments. This reinforces the investments on this area and justifies the value of the presented work that pretends to expose a methodology to identify components without using markers.

In order to set up the 3D vision system correctly, it is also presented interfaces to generate calibration plates for different scenarios and to acquire images to calibrate the whole system, through user-friendly interfaces.

2. STATE OF THE ART

The AR is based on a vision system to capture images to be processed in real time. This processing involves overlaying virtual data on the real scenario captured by the vision system and allowing interaction with the result of fusion between real and virtual environments, [2]. Other definitions come from [3] and [4] where an AR application should fulfil three requirements, or be classified within the "Reality-Virtuality continuum", [5]-[6].

An AR application gives the possibility to see invisible things to the human eyes. Through virtualizations it is possible to display the flow of heat dissipated on equipment or even the location of electrical wires on a wall. These are just a few examples where AR may be an important aid to conduct, analyse and help the technician to intervene in technical assets and also for project phases to display the results of changing certain characteristics in the project.

Despite the efforts to apply this technology on industries, several prototypes were tested on different areas like the following ones: to support assembly process of medical (surgical) equipment, [7]; as surgical support tool to classify the organs, [8]-[10]; as navigation system on nuclear power plants, issuing warnings about the dangers along the paths, [11]; to present advanced technical instructions and demonstrating how to assembly parts, [12]; to support military maintenance interventions, [13]. All these projects have been developed under several demands, related to the critical requirements of those sectors.

There were also other projects, developed among different European countries, [14]-[16], which aimed to create tools or reach methodologies to be applied on the maintenance sector in which were tested new AR devices with the objective to evaluate their benefits and find the gaps to be filled in new projects.

Despite the problems detected and their restricted functionalities, all of those projects reached promising results, mainly associated to minimization of committed errors and reduction of the intervention times, what means, in general terms, improvement of maintenance efficiency.

The benefits of AR are mostly related to the identification of components on the field of vision of the technician and how virtually data is overlapped. Usually, the service manual shows how to proceed, step-by-step, through illustrations and textual information. The same can be done through AR directly over target components in captured scenario, presenting step-by-step the procedures of how to assembly or disassembly parts correctly. In order to create an advanced AR tool to support maintenance interventions, this must be completed with more functionalities, in relationship to previous prototypes, as is referred in section 6.

Nowadays, in industry, many of assembly tasks are automated but other ones still requiring human labour being some of them dangerous to human health. It is for these tasks, which need human intervention that AR is able to contribute to achieve better efficiency on maintenance and minimize human risks by introducing remote assistance with AR. According to tests that have been done, an application based on AR can improve the final result reducing the error rate up to 82%, [17].

Voice recognition is an extra functionality that helps to improve the benefits of applying AR on maintenance tasks because it allows a quick access to additional data of the component, [18], and promote the productivity by displaying all the information on an intuitive way, [19].

The major gap of most projects already developed is about the methodology used to identify the components, which has been made through markers that are used due to constraints in the identification of components by their geometry on industrial environments. The use of markers leads to another tricky situation that is related to the necessity of marking all components that

need to be identified, what it is extremely difficult to implement. Additionally, the major number of AR applications requires that the marker remains always detectable during interventions, what is a strong operational limitation. When this not occurs the AR tool can suffer interferences that cause negative effects on the performance of technician, as yet experienced, [7].

To identify parts on real scenario it is proposed a multi-camera vision system, with the objective of testing two different methodologies, one that generates a 3D point cloud of the captured scenario and try to find the target component on it, and the other one based on the shape of components to be detected.

The vision systems are traditionally used as a quality control tool in industry. These systems have as advantage the capacity of analysing a lot of information, with high level of detail, in less time when compared with humans. On manufacturing lines a vision system can be combined with a light system or even added an environment controlled chamber to achieve the perfect conditions to capture image or reconstruct the scenario. The reconstruction is often made by a camera and a laser, or recently by Time-of-Flight (ToF) cameras. The ToF cameras are recent on the market and for some applications they aren't good enough yet because their low pixel resolution and the excessive noise on depth readings, [20]-[21].

On multi-cameras systems is also important the relative positioning between optical-axis of cameras, which should be adjusted accordingly to the target function, as exposed in [22].

3. VISION SYSTEM AND PROCEDURES

Having as objective to develop a robust vision system that can identify and track components by their shape, i.e., without using markers, it is under research a methodology that better suits this goal. During this research it is also identified which method is better for different application areas, like workbench works or a portable vision system hold by the technician.

Apart from application area, the system should retrieve accurate pose results, even when partially occlusions occur, to overlay correctly the virtual data on the field of vision of the user.

To meet the objectives it is used a multi-camera system, that is a requirement for 3D reconstruction to match the same point between pairs of cameras in order to get their spatial coordinates and for the method centred on the shapes of components, the second camera is used to validate the correct identification of target component made by the first camera.

A 3D vision system needs to be calibrated to obtain correctly dimensions and position of parts and even the cameras pose, which will be needed to create ROI's (Region Of Interest) on the image of second camera based on component positioning in the first camera.

As software to processing images it was chosen the Halcon, from MVTec, due to their capability to work on 3D vision and to use the CAD model of components to generate the parameters needed to identify them on the captured scenario. This software has his own calibration plate model, but it is possible to calibrate the system with other calibration plates models or the same model with non-standard dimensions.

The calibration plate should be chosen accordingly to the environment characteristics and the estimate dimensions of

captured area, what means that for some situations a non-standard calibration plate has a high probability to produce better results. Based on this statement it was created an interface, with an API (Application Programming Interface) to CAD software that generates the CAD model of the calibration plate and their parameters file needed to calibrate the vision system with Halcon, using the dimensions set by the user. With this simple interface the user won't need knowledge on CAD drawing software; he/she just needs to set the dimensions of some characteristics of the desired calibration plate.

To help the user to choose the right dimensions for the calibration plate, this interface also propose them based on, the size of the working area to be captured, the distance between cameras, the working area, and cameras characteristics.

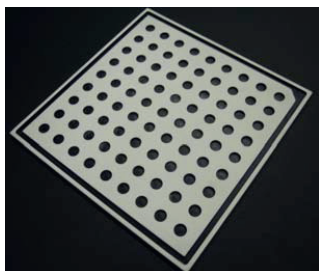


Figure 1. Customized manufactured calibration plate

From CAD model of the calibration plate, it can be easily manufactured by a CNC milling machine, ensuring the exactly dimensions according to the parameters file. On Figure 1 it is presented a manufactured calibration plate of 166 mm side. Some tests were done with one customized manufactured calibration plate that warrants their use to correctly calibrate the vision system. This interface represents a solution that allows excellent flexibility at low-cost.

To locate and get the correct coordinates of components, it is demanded to calibrate the vision system. To do that operation it was built other interfaces to acquire images and calibrate the system based on those images. Those interfaces are simple to manage, but the calibration process requires basic knowledge on image processing to set up all configurations.

After selecting an appropriate calibration plate, acquiring images of that plate and calibrating the system, it is possible to proceed to the components detection stage.

3.1 3D Reconstruction

The first attempt to find a component is through the reconstruction of the real environment using, at least, two cameras, representing the captured scenario on a point cloud. One of the biggest constraints of this method is the correlation of points between two cameras, needed to estimate the deep of a point.

Without using lasers or ToF cameras, the surface of components shouldn't be smooth, because this represents an harder task to get the correlation of points between cameras that usually fails; by other words, it can't retrieve the three coordinates of a point and can't represent it on a point cloud.

As exposed behind, the success of this method is related to the surface of components, but this isn't the unique characteristic that influences the process - the cameras resolution and their positioning also influences it. Cameras resolution is linked to more or less quantity of pixels for the same captured area, while the cameras positioning is associated to the common captured area and features.

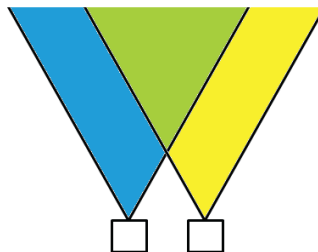


Figure 2. Parallel optical-axis model

In figures 2 and 3 are represented two optical-axis models in which the blue and yellow areas represent the areas captured by the left and right cameras, respectively, while the green area represents the common captured area on both cameras. On parallel optical-axis (Figure 2), it can be tried to decrease the distance between cameras in order to enlarge the common area - the distance between cameras, also known as baseline, must be chosen according to estimated cameras to object distance and characteristics of the object [23] - typically it is recommended to increase the baseline in both models.

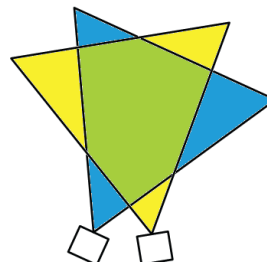


Figure 3. Converging optical-axis model

The components to be reconstructed must be in the common area viewed by cameras, because only the components in that area are theoretically captured by both cameras – as said before, both projections are required to retrieve the spatial coordinates. It is due to occlusions or shadows of some components that can't be captured, at least by one the pair of cameras.

Tests done with different components and diverse geometries, [24], revealed that converging optical-axis model retrieves a higher measuring accuracy.

As mentioned before, objects surface must have texture; otherwise, the point's coordinates can't be determined because

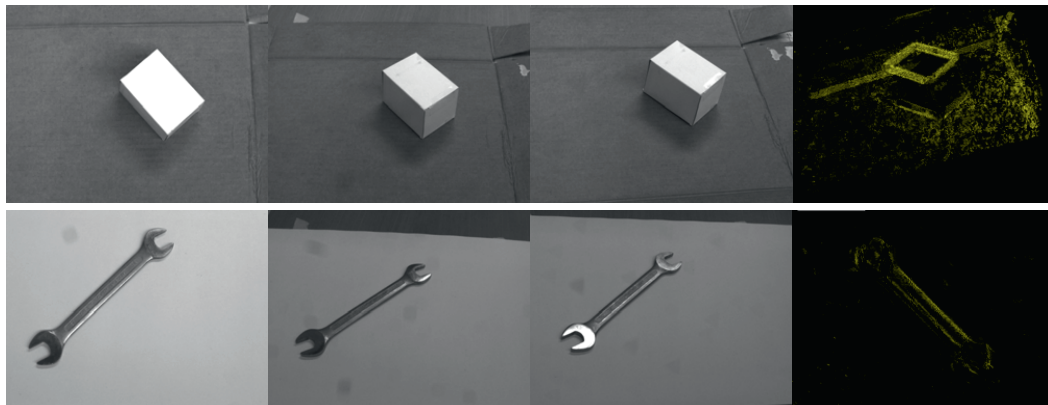


Figure 4. Scene reconstruction - most right images represents the point cloud from combination of others three images

the correlation of points is made through the availability of texture, [23]. Too large distances between cameras can result as a failure on correlation of features between cameras, as illustrated on Figure 5, where just a small component part is common to both cameras - the others are naturally occluded due to object geometry and cameras positioning.

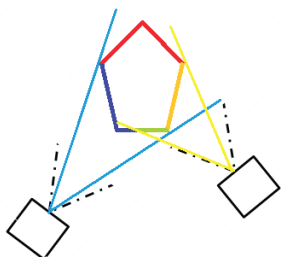


Figure 5. Cameras field of view of an object

Based on performed tests it was verified that, on a system with three cameras, one camera must be placed to acquire images from top while the others should acquire images from side view to achieve better results on generation of point cloud.

The use of three cameras, with one from top view, has natural restriction for portable solutions, but it can be successfully applied on workbench works if the process, since image acquisitions until identifying target components on scene - is accurate enough and fast to be considered as real time.

To evaluate the performance of reconstruction and the process of identifying parts in a cloud of point, it was used three cameras with resolution of 656x494 pixels, and acquired several images of the objects from different positions. Figure 4 shows the three images acquired by each cameras and the last one represents the cloud of points. The represented cloud of point displays a poor reconstruction due to lack of texture on the surface of the objects. It was also applied a sharpening filter to try getting better results but reconstruction remained with poor quality.

This unsuccessful attempt to represent in a point cloud the captured scene prevents the test to identify components. It is known that it will be harder to reconstruct the real environment, with poor texture, only by cameras, in an open field, without a proper light system, but at least it was possible to identify the best combination of cameras to use.

As the captured area was 200x265 mm, they were made the same tests with higher resolution cameras, with one megapixel, which retrieves better results on reconstruction but still fail on identification of components. This failure on the identification is due to lack of texture on surfaces - the system reconstructs only the objects contours.

Besides a poor reconstruction and the difficulties of the system to identify the parts on most cases, for the human eye it is perfectly possible to identify those parts in a point cloud through their contours. The time consumed in reconstruction process is excessive for a real time application.

3.2 Identification by shapes

On this alternative method that identifies the components by their shapes, it is stored the projections of the target according to predefined conditions, such as: distance range between camera and object; image pyramid levels; and angles at which the object will be seen. These conditions must be correctly defined in order to avoid overdata storage.

The first attempt was to identify the target parts directly over the acquired image from a single camera. On this test it was reached a successful rate of 71% where any light change requires adjustments of the configurations, due to direct use of acquired image. The optimum situation is to apply the detection method on images without noise and with well-defined components contours, but this is impossible to get on in an open field application even the objects shadows cause occlusions to the vision system. Considering this and aiming to develop a more flexible solution and to improve the successful rates, it was tested different combinations of filters to highlight objects contours. Applying those it was possible to achieve different

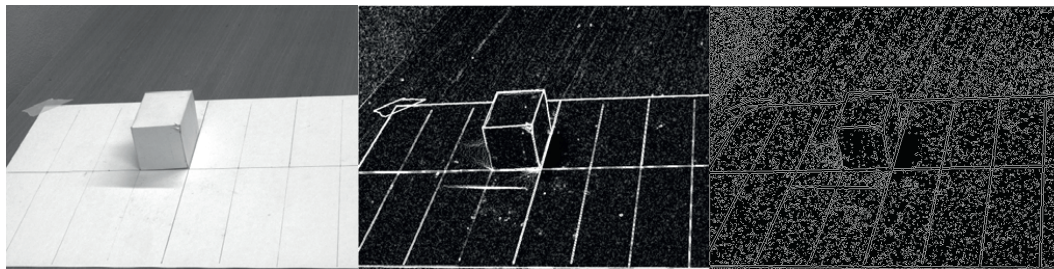


Figure 6. Edge detection, from left to right: a) acquired image; b) filtered image; c) Halcon internal pre-processed image over b)

successful rates, depending on the combination that was used, since 64% until 84%.

Although the filtered images enhance the successful rate on detecting parts, these images remains very noise (Figure 6 b) that could cause failures.

The process to identify parts in the captured scenario, by Halcon, also made some image processing internally that cannot be avoid, Figure 6 c). Besides image Figure 6 b) looks fine to run the identification process, the image Figure 6 c) shows that remains excessive noise.

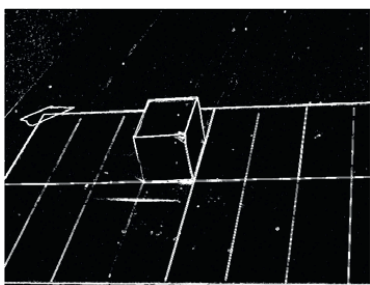


Figure 7. Image filtering applied over Figure 6 a)

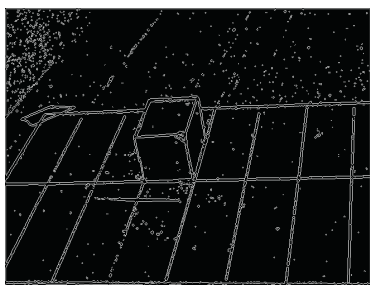


Figure 8. Halcon internal pre-processing over Figure 7

To remove that noise it was developed a new filtering methodology that applies not only sharpening filters but also smoothing filters and histogram based image processing. The results of this methodology to remove noise are clearly better than the previous, as illustrated in figures 7 and 8.

This method, based on shape recognition, doesn't need a stereo vision system - it works only with one camera - it can't distinguish parts with the same shape on different scale.

To run this method it was used tests to setup with two cameras, where the second camera was used as a backup camera, to classify the detection made by the first camera as acceptable or not.

Over the image of the second camera it is also applied the same filtering methodology developed on just a small region of the entire image. A small region is applied a ROI in order to restrict the area to be scanned, what allows to avoid unnecessary time consumption. The ROI was created based on the detected position of the object from the first camera, to transpose their coordinates to the image of the second camera.

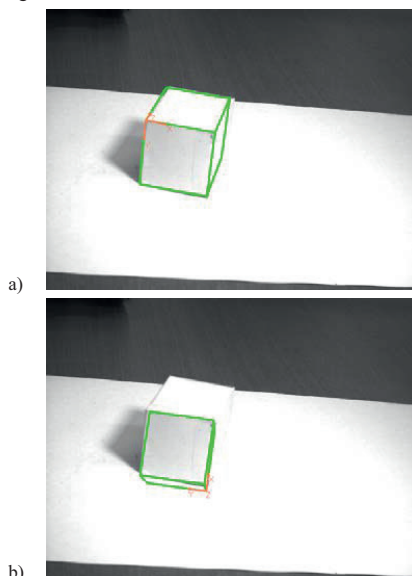


Figure 9. Part detection applying filters a); without filters b)

The filtering methodology developed, also allows the correct identification of the target component on a low contrast image, as represented in Figure 9.

It is intended to extend this filtering methodology to robots, particularly in bin-picking tasks, not only due to their scientific interest, [25], but also because, when applied in industry, can prevent injuries by doing automatically and on autonomous way the dangerous tasks.

4. PROPOSED AR SYSTEM

The complete system is composed by an AR tool that guides the technician during the intervention which is connected to an Expert System (ES) and an information system like SMIT **Error! Reference source not found.** The importance of incorporate an ES is to guide the technician solving faults through an interactive smart natural voice interface. The connections with SMIT, or similar systems, is helpful because the historic of equipment permits that the technician will be able to interact with the information system and the ES to get additional support to aid decision for the solution of the faults, what permits to optimize time and improve the interventions quality.

The AR system has as base a vision system that has to be able to identify the components through their geometry to ban the use of markers, which is a non-valid option, as represented at Figure 10. To aim that goal it can be applied a solution like the presented previously on section 3.2.

The proposed AR system applied on assembly tasks on a workbench may have the capability to inform the list of required tools for a job and checking if all the components needed for an assembly task are available on the workbench. Each assembly step will be guided by the AR system, showing how to perform every steps and giving warnings to prevent mistakes. Especially for the trainees, additional information can be provided, like how to handle components and evaluate their performance on real time.

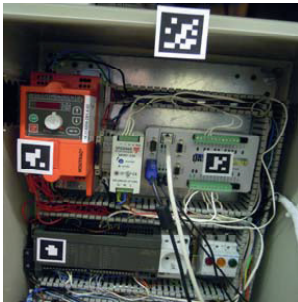


Figure 10. AR markers

Some maintenance tasks aren't common and even the best technician can have doubts about how to proceed on special occasions. For these situations, the AR system may be equipped with tools that allow remote assistance. Through this method a specialist may intervene in the field view of the technician, by virtual data to display more information in order to identify components manually or even to show how to handle instruments in order to aid the technician on the field. Remote assistance requires more human resources; additionally it can be a simplified and effective way to finish the job, even if the technician is far away from headquarters. This also requires that technicians use some equipment to capture the real environment and to receive

the virtual information on a display, which can be done through AR glasses, tablets or smartphones through connection to the internet.

The remote assistance can also be applied to remote interventions by robots, where the user can control the robot through gestures [26]. The AR will be used to identify the target parts on the scene and to display additional information about the scene and thereby avoid exposing humans to hazardous environments.

In case of equipment failure during the warranty period, if the failure was simple to solve, can be used an AR based application that demonstrate the owner how to repair the equipment. On this case, the AR must evaluate the performance and warning the user to avoid faults. For the owner this means less time without equipment and the manufacturer can save time that will be spend by human resources to solve simple tasks to do other tasks.

5. CONCLUSIONS

A complete AR tool for industrial maintenance, as presented in this paper, shouldn't be available on short term, due to technological restrictions and to the demands of industrial sector. It is necessary to test new prototypes to evaluate the proposed functionalities and thus enhancing its efficiency for maintenance field in a near future.

Besides other projects that have their focus on how to display virtual data, it is also extremely important, as shown in this paper, to find a way to avoid markers, detecting the components by their shape and through user friendly interfaces.

Based on the achieved results, the 3D reconstruction method used to identify components in a point cloud, only by cameras, isn't a choice and the ToF cameras remains with a lot of noise that must be processed; however, they will be a valid solution after some improvements. On the other hand, the methodology used to identify parts based on their shapes represents a valid solution to be applied since it is robust enough to occlusions.

A robust vision system that is able to identify components without using markers is one of the main characteristics for an AR system which, when applied to the maintenance sector, will improve the performance of technicians during maintenance interventions and minimize downtimes and costs. All of this contributes to the increasing of the efficiency and quality of maintenance interventions.

6. REFERENCES

- [1] A. Nee; S. Ong; G. Chryssolouris; D. Mourtzis (2012): *Augmented reality applications in design and manufacturing*. CIRP Annals - Manufacturing Technology, Vol. 61, No 2, pp. 657-679, ISSN 0007-8506.
- [2] R. Oliveira; T. Farinha (2012): *Realidade Aumentada – Ferramenta de Apoio à Manutenção em Equipamento Hospitalar*. Tecno Hospital, No. 49, pp. 12-16.
- [3] R. Azuma; (1997): *A Survey of Augmented Reality*. Presence: Teleoperators and Virtual Environments, Vol. 6, No. 4, August, pp. 355-385.
- [4] R. Azuma; Y. Baillot; R. Behringer; S. Feiner; S. Julier; B. MacIntyre (2001): *Recent Advances in Augmented Reality*. IEEE Computer Graphics and Applications, Vol. 21, No. 6, pp. 34-47.

- [5] P. Milgram; F. Kishino (1994): *A Taxonomy of Mixed Reality Virtual Displays*. IEICE Transactions on Information and Systems, Vol. E77-D, No. 12, pp. 1321-1329, December.
- [6] P. Milgram; H. Takemura; A. Utsumi; F. Kishino (1994): *Augmented Reality: A Class of Displays on the Reality-Virtuality Continuum*. Proceedings of Telem manipulator and Telepresence Technologies (SPIE), Vol 2351, No 34, pp. 282-292.
- [7] S. Nilsson; B. Johansson (2007): *Fun and usable: augmented reality instructions in a hospital setting*. Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces (OZCHI '07), ACM, pp. 123-130, New York, NY, USA.
- [8] B. Reitinger; P. Werlberger; A. Bornik; R. Beichel; D. Schmalstieg (2005): *Spatial measurements for medical augmented reality*. Proceedings of Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 208-209, 5-8 October.
- [9] K. Gavaghan; M. Peterhans; T. Oliveira-Santos; S. Weber (2011): *A Portable Image Overlay Projection Device for Computer-Aided Open Liver Surgery*. Biomedical Engineering, IEEE Transactions on , Vol 58, No 6, pp.1855-1864, June.
- [10] H. Jeffrey; M. Shuhaiber (2004): *Augmented Reality in Surgery*. Arch Surg, Vol. 139, pp 170-174, February.
- [11] H. Ishii; Z. Bian; H. Fujino; T. Sekiyama; T. Nakai; A. Okamoto; H. Shimoda; M. Izumi; Y. Kanehira; Y. Morishita (2007): *Augmented Reality Applications for Nuclear Power Plant Maintenance Work*. Proceedings on CD-ROM of the International Symposium on Symbiotic Nuclear Power Systems (ISSNP) for 21st Century, pp.262-268.
- [12] A. Sá; G. Zachmann (1999): *Virtual Reality as a Tool for Verification of Assembly and Maintenance Processes*. Computers & Graphics, Elsevier, Vol 23, No 3, pp. 389-403.
- [13] S. Henderson; S. Feiner (2009): *Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret*. Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality (ISMAR '09), IEEE Computer Society, pp. 135-144, Washington, DC, USA.
- [14] ARTESAS project, <http://www.artesas.de>
- [15] J. Weidenhausen; C. Knoepfle; D. Stricker (2003): *Lessons learned on the way to industrial augmented reality applications, a retrospective on ARVIKA*. Computers and Graphics, Vol. 27, No 6, pp. 887-89, ISSN 0097-8493, December.
- [16] B. Schwald; B. de Laval (2003): *An Augmented Reality System for Training and Assistance to Maintenance in the Industrial Context*. Journal of WSCG, Vol 11, No 1, ISSN 1213-6972, Plzen, Czech Republic, 3-7 February.
- [17] A. Tang; C. Owen; F. Biocca; W. Mou (2003): *Comparative effectiveness of augmented reality in object assembly*. In Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '03), ACM, pp. 73-80, New York, NY, USA.
- [18] D. Espindola; C. Pereira; M. Pinho (2008): *Visualization of Intelligent Maintenance Systems Through Mixed Reality*. Proceedings of VRCAI '08 Proceedings of The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, ACM, Vol 1, No 212, pp. 5-6, New York, NY, USA.
- [19] H. Lipson; M. Shpitalni; F. Kimura; I. Goncharenko (1998): *Online Product Maintenance by Web-Based Augmented Reality*. Proceedings of CIRP Design Seminar on New Tools and Workflows for Product Development, pp. 131-143.
- [20] S. Gudmundsson; H. Aanaes; R. Larsen (2008): *Fusion of stereo vision and Time-of-Flight imaging for improved 3D estimation*. Int. J. Intell. Syst. Technol. Appl., Vol. 5, No. 3/4, pp. 425-433, November.
- [21] Y. Kim; D. Chan; C. Theobalt; S. Thrun (2008): *Design and Calibration of a Multi-view TOF Sensor Fusion System*. Proc. IEEE CVPR Workshops, Vol. 1-3, pp. 1524-1530, June, Anchorage.
- [22] J. Cardenas-Garcia; H. Yao; S. Zheng (1995): *3D reconstruction of objects using stereo imaging*. Optics and Lasers in Engineering, Vol. 22, No. 3, pp. 193-213, ISSN 0143-8166
- [23] Halcon MVTEC (2011): *Solution Guide III-C 3D Vision*. Ed. 5a, April.
- [24] J. Cardenas-Garcia; H. Yao; S. Zheng (1995): *3D reconstruction of objects using stereo imaging*. Optics and Lasers in Engineering, Vol 22, No 3, pp. 193-213, ISSN 0143-81
- [25] S. Fuchs; S. Haddadin; M. Keller; S. Parusel; A. Kolb; M. Suppa (2010): *Cooperative bin-picking with Time-of-Flight camera and impedance controlled DLR lightweight robot III*. International Conference on Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ, pp.4862-4867, 18-22 October.
- [26] José Torres Farinha, Inácio Fonseca, António Simões, Fernando Maciel Barbosa, José Viegas (2008): *New ways for terology through predictive maintenance in an environmental perspective*. WSEAS Transactions on Circuits and Systems, Issue 7, Volume 7, July 2008. ISSN 1109-2734, pp630-647. (The ACM Digital Library).
- [27] P. Neto; J. N. Pires; A. P. Moreira (2010): *High-level programming and control for industrial robotics: using a hand-held accelerometer-based input device for gesture and posture recognition*. Industrial Robot-An International Journal, Vol 37, No 2, pp. 137-147.

Current and Prospective Information and Communication Technologies for the E-maintenance Applications

Jaime Campos
Linnaeus University, Sweden
Department of Informatics
jaime.campos@lnu.se

ABSTRACT

This paper is about the current state of affairs when it comes to the Information and Communication Technologies (ICTs) in Condition Monitoring (CM) and Maintenance. In addition, there are significant efforts done to standardise the systems for CM and Maintenance under discussion. The ICTs such as the Web technologies are presented, since they are an integral part of the E-maintenance approach. The author goes through the latest developments in the area of Web technologies and the studies conducted by the World Wide Web Consortium (W3C). It emphasises the Semantic Web and its standardizing practices, the Web 2.0 and its Social media technologies, which are non-existing to the moment at the e-maintenance applications. It discusses applications in the domain of interest and how the latest ICTs, especially the developments of the Web technologies and the Cloud computing, can make impact and might affect the future e-maintenance applications.

Keywords: Web technologies, Web 2.0, Social media technologies, e-maintenance, Cloud Computing.

1. INTRODUCTION

It is well known that the technologies such as the Information and Communication Technologies (ICTs) are used to support various business processes and if they are as well planned, they even support the business strategy of an enterprise. Even if they are just a support tool for business, it's important to understand their current state, characteristics and evolution. In addition, it is important to understand how companies can benefit from this development, which is a reality for all companies using ICTs in their organization to support different processes. This enables to use them in an optimal manner for purposes that benefits the whole enterprise. In maintenance, the application of software and computers got an impetus with the coming up of desk tops. Initially the software was used for record keeping and job scheduling etc. In 1980s AI techniques, like expert system, started appearing for the use in condition monitoring. In 1990s, papers started coming about the application of other techniques like Neural Network and Fuzzy Logic in condition monitoring. Reviews on e AI techniques [1, 2, 3, 4] are available. Distributed artificial intelligence (DAI) and intelligent agent technology were next to develop [5]. Applications of DAI were also done to condition monitoring [6, 7, 8, 9] but agent technology, in general, got wider acceptance [5]. For some years, attempts have been made to introduce standards in software industry. Such attempts have the advantages of easier upgrading of system parts, more suppliers systems, providing more alternatives for faster technology development and reduced prices [10]. However, the

pace at which developments are occurring in the software industry, inventions of new languages and protocols, and the standardisation appears to be difficult in the near future. Simultaneously, there have been attempts also to standardise the systems and subsystems for condition monitoring and maintenance. One such attempt is OSA-CBM.

The purpose of this paper is to briefly go through the development of the ICTs in the area of Condition monitoring and Maintenance, the e-maintenance approach and the emergent ICT's, especially the Web technologies such as the Web 2.0 and Cloud computing and highlight its characteristics and discuss how they might affect the maintenance department, for instance when the e-maintenance, Web 2.0 and Cloud computing are applied. Further on the efforts of the OSA-CBM are highlighted.

2. OPEN SYSTEM ARCHITECTURE CONDITION BASED MAINTENANCE

Condition Based Maintenance (CBM) is becoming more and more recognised in the U.S. industry and the military. CBM system contains a number of features. The system function requires the integration of various hardware and software components. These components can come from several sources, but it should be possible to integrate and share them. Open System Architecture and Condition Based Maintenance (OSA-CBM) started as a solution to this problem. The aim was to promote the integration and interoperability of components available from different sources. They work to create de facto standards, which cover all the features of data collection on the recommendation of specific maintenance tasks.

OSA-CBM was an industry initiative, partly funded by the US Navy through a DUST (Dual Use Science and Technology) to develop and demonstrate an Open System Architecture for Condition Based Maintenance. Participants come from industrial, commercial and military CBM technology. Other important contributors were the MIMOSA CRIS and IEEE standards. With Internet and LAN, (local area network) distributed software architectures can be used for CBM due to the fact that it is cost-effective. The architecture of OSA-CBM has 7 layers [10] (www.osasbm.org). These are the following: Layer 1- Data Acquisition: It senses the signal from the machine and digitizes it. The digitized sensor data is then passed on to the second layer. Layer 2- Data Manipulation: In this layer, different signal processing operations are performed on the data received from the sensor and/or other signal processing modules. Layer 3- Condition monitor: It compares the received values from the previous layer with the expected values. It may also generate alerts or give the operational state of the machine. Layer 4- Health assessment: It receives data from condition

monitor or other health assessment modules. The main function of the module is to determine if a deterioration has occurred in the machine and if it has, it is expected to provide the different relevant fault conditions. Layer 5- Prognostics: The main function of this layer is to give a projection on the future health of the machine or its remaining useful life (RUL). While doing so, one should consider different estimates of future usage profiles of the machine. Layer 6- Decision support: The primary function of this layer is to recommend actions and in case of multiple alternatives, the implications of each alternative. One of the alternatives, for example, can be the maintenance action schedule. Layer 7- Human Interface (Presentation layer): This layer provides displays health of the machine (layer-4), prognosis of its condition (layer-5) and recommendations for action, if any, (layer-6) to the user. When an abnormal condition is reported, the user can dig into multiple layers to trace the sources of data and the process of the decision making.

In 1994 a non-profit organization founded the Machinery Information Management Open Systems Alliance (MIMOSA). The aim was to promote the exchange of information among the plants on machine maintenance, to come up with publications and develop open international agreements (standards) on machine maintenance information systems, [10] (www.mimosa.org) MIMOSA developed a Common Relational Information Schema (CRIS). It is a relational database model for different data types that need to be processed in CBM application. The system interfaces have been defined according to the database schema based on CRIS. The interfaces' definitions developed by MIMOSA are an open data exchange convention to use for data sharing in today's CBM systems. There are several condition monitoring and maintenance systems such as the PROTEUS, TELMA AND CASIP etc. these and several other systems are gone through in Muller et al. [11]. However, for the last two decades, artificial intelligence has been applied to CM and maintenance. Some literature available on them is [1, 2, 3, 4]. The first review on the subject, the developments of ICT in CM and Maintenance, was published in 2006 [12] and an extended and updated version was published in 2009 [13].

First and foremost step is to study Web technologies for condition monitoring and maintenance and then to distinguish e-maintenance. In the next section the e-maintenance concept is presented. Thereafter the Web technologies and Cloud computing are gone through.

3. E-MAINTENANCE

The e-maintenance is a multi-faceted approach. It is more than implementing a maintenance strategy, maintenance plan or a maintenance type [14]. Through the framework proposed by Lung et al. [15] and Levrat et al [16] it is possible to get an insight into the comprehensiveness/broad range of the e-maintenance concept. This is especially possible when observing the abstraction levels proposed by the authors. The abstraction levels are the e-maintenance strategic vision, e-maintenance business processes, e-maintenance organization, service and data architecture, and e-maintenance IT infrastructure. However, for the successful implementation of the e-maintenance concept the complete integration of data, system and processes is an important requirement when the assets are geographically distributed. ICTs can play a vital role in achieving this objective. According to Lung

[17] e-maintenance is a sub discipline of e-manufacturing and e-business for the support of next generation manufacturing practices. Accordingly, he defines e-maintenance as "the ability to monitor plant floor assets, link the production and maintenance operation systems, collect feedbacks from remote customer sites, and integrate its upper level enterprise applications". A more general definition is "maintenance management concept whereby assets are monitored and managed over the internet". Thus, the author mentions that this development is the cause of the advances on the new technologies, like; Internet, web and wireless etc., which have led to the replacement of the conventional reactive strategies by proactive i.e., aggressive strategies. In conclusion, Lung [17] believes that it is a revolutionary change rather than an evolutionary advancement. Muller et al. [11] mention that this development takes us to a new maintenance generation, which has emerged since early 2000. There are many efforts done in the e-maintenance direction both in academia and industry, examples of these can be found in, Holmberg, et al. [18]; Muller et al. [11]; Karim, [19]; Campos Jeria [21].

The next ICTs illustrated are the ones supporting the e-maintenance, i.e. the Web technologies and the emerging ICTs such as the Web 2.0 and Cloud computing, their role in e-maintenance in the infancy phases.

4. WEB TECHNOLOGIES

Through Web technology, the information available on the net can generally be shared by all (Internet) unless restricted by design (intranet). This is possible through a common framework such as the Semantic framework. It provides the data to be shared and reused across various applications and organisations (www.w3.org). This framework is a common result from a large amount of researchers and industrial partners led by W3C. The Semantic web framework uses common format for integration and combination of data drawn from different resources. It provides Web pages, databases, services, programs, sensors and personal devices to use and produce data on the Web [21]. The Semantic Web consists of technologies such as Unicode, Uniform Resource Identifier (URI), Extensible Markup Language (XML), XML Schema, Resource Description Framework (RDF) Core Model, RDF Schema language, Web Ontology language (OWL), Logic, Proof, and Trust (www.org). Unicode and URI layers ensure that the characters used are identified in the Semantic Web. The XML layer consists of name spaces and schema definitions. The objective of this layer is to support the data exchange for any application developed by any platform, i.e. involving different hardware, operating systems and programming languages. The RDF core model and RDF schema facilitate to make statements about objects and define vocabularies that can be connected to URI. The RDF is a recommendation with the objective to standardize the use of definitions and metadata, i.e. presentation of data, information, knowledge, or resources located on the Web. The Ontology vocabulary defines relations among different concepts. Digital signature layer provides support for sensing updates to documents. The other parts of the Semantic Web, i.e., Logic, Proof and Trust are in the experimental phase. It exists only in simple application prototypes that are being developed. Logic layer writes the rules, Proof executes these rules, and Trust evaluates them in a validation process. Except these three, all the other layers have been standardized.

The Semantic Web framework meaning has to do with the W3C, significance or vision, which consists of web linked data. The framework allows people to create data storage on the Web, to build vocabulary and to write rules for data management (w3.org). This is possible and empowered by technologies such as RDF, SPARQL, OWL and SKOS.

4.1 Web services

The World Wide Web Consortium (W3C) defines the Web Services (WS) as being software designed to support interoperability between distributed applications over a network (www.w3.org). The WS consists of technologies such as HTTP, XML, SOAP, WSDL, SPARQL, and others. The SOAP protocol uses HTTP to transport its messages. The WS provides the system with integration possibilities. Java Sun explains WS as being an application that uses open, XML standards, which use transport protocols to exchange the data with its various clients (java.sun.com). The WS consists basically of three components. These are the XML, utilised in the different layers of the WS, the second is a soap listener, which packages, sends and receives messages over the Internet with Hypertext Transfer Protocol (HTTP) and Transmission Control Protocol/Internet Protocol (TCP/IP). The code that the client machine uses to read the WS messages is the described by Web Services Description Language (WSDL), which is the third component.

A WS-agent is a software that acts on behalf of a person or organization. The WS agent can request and perform zero or more services (www.w3.org).). WS is an abstract notion that must be implemented by a concrete agent. The agent sends and receives messages and becomes in this way a means to offer and provide the WS. The provider entity, in this case a person or organization, permits an agent to perform a particular service and a requester entity, i.e. a person or organization uses the WS. This is done through a requester agent, which facilitates the exchange of messages with the provider agent. In most of the cases the requester agent initiates the exchange. When a WS is requested, the first step for the requester and the provider agent is to get to know each other. The second step lies in the requester and the provider entities, i.e. agents agree on the description and semantics to be used in their interaction. The third step suggests that they should use the descriptions and semantics. Finally, the fourth step presupposes interaction between the requester and the provider entities (agents), i.e. they send and exchange their messages on the enquiry.

4.2 The Web 2.0

The Web 2.0, even called Social Media Technologies, involves the emergence of technologies such as AJAX, flash, social network, Semantic Web with its WS and ontologies etc. [22]. It is defined as a technology that “emphasizes the value of user-generated content. It is about sharing and about communication and it opens the long tail which allows small groups of individuals to benefit from key pieces of the platform while fulfilling their own needs” [23]. Another definition of Boateng, et al. [24] is that the “Web 2.0 facilitates knowledge creation and sharing by involving, engaging and empowering people and by creating a collaborative environment for social interaction between those who need to seek knowledge and those who hold the knowledge”.

Moreover, McGee & Begg [25] defines it as “Web 2.0” describes a collection of web-based technologies which share a user-focused approach to design and functionality, where users actively participate in content creation and editing through open collaboration between members of communities of practice”.

Further, the Web 2.0 applications implicate new web content types, as well as new level of user interface design, i.e., the user interfaces are richer as well as the user generated content and the level of interface design in Web development. The main characteristics of the Web 2.0 and its tools lie in giving possibilities to collaborate, create network, communicate and share information as well as create network and scale effects in grand communities. The possibilities to be an instant publisher enable at the same time to be an author and publisher. The companies that implement the Web 2.0 need to acquire organizational capabilities to be able to react to the user-employee-generated content as well as to the interaction among various employees. Therefore, companies that are currently using Web 2.0 technologies create platforms that facilitate the open collaboration and simultaneously give exchange of information and knowledge internally in the company, even between customers and suppliers. The Web 2.0 technologies, such as for example Wiki' and RSS, offer other possibilities apart from Web 1.0, due to the fact that the content can be created by all employees in the company and not only by the programmers. The information is more dynamic, and the Web pages consist also of applications. Their architectures are based on WS. The new possibilities, which the Web 2.0 provides on Condition Monitoring and Maintenance as well as e-maintenance, are presented at Campos & Jantunen, [26]. In conclusion, the Web 2.0 supports even a learning environment that companies need to investigate on how to use their specific needs and domain for the interest of the whole enterprise [27].

5. CLOUD COMPUTING

The Cloud computing emerged during the late 2007 and its main thrust is to provide on-demand computing services with high reliability, scalability and availability in a distributed environment [28]. It is defined, by the National Institute of Standards and Technology (NIST), (www.nist.gov), as “a model for enabling ubiquitous, 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.” Another definition, by Wang & Laszewski, [29], is the following: “A computing Cloud is a set of network enabled services, providing scalable, Quality of Service guaranteed, normally personalized, inexpensive computing platforms on demand, which could be accessed in a simple and pervasive way.”

Another definition by Marston et al. [30], which does not require that the services are given by a third party is as follows: “It is an information technology service model where computing services (both hardware and software) are delivered on-demand to customers over a network in a self-service fashion, independent of device and location. The resources required to provide the requisite quality-of service levels are shared, dynamically scalable, rapidly provisioned, virtualized and released with

minimal service provider interaction. Users pay for the service as an operating expense without incurring any significant initial capital expenditure, with the Cloud services employing a metering system that divides the computing resource in appropriate blocks.”

In Cloud computing is everything seen as a service, i.e. Software as a service (SaaS), Platform as a service (PaaS), and Infrastructure as a service (IaaS) [28]. Authors, such as Armbrust et. al. [31], refer Cloud Computing to both the applications delivered as services over the Internet and the hardware and systems software in the datacentres that provide those services. The services themselves have long been referred to as Software as a Service (SaaS). Such services as for instance Software as a Service, (SaaS) mean the access to some software applications through the internet, and at the same time one can avoid the “binding” to software and hardware management [32]. The software can range from word processing programs to ERP applications. The Platform as a Service (PaaS) involves the network, database and system management experts, which with their expertise keep and maintain the operating systems, databases, web servers, etc., which are at the Cloud. The Infrastructure as a Service (IaaS) concerns the possibility to access through the Internet servers, storage devices, etc.

However, the advantages of SaaS to both users and service providers are well understood. Service providers enjoy greatly simplified software installation and maintenance and centralized control over versioning; and users in their turn can have access to the service “anytime, anywhere”, share data, collaborate more easily, and keep their data stored safely in the infrastructure [32]. Furthermore, the authors mention some new specific characteristics of the Cloud computing, especially what is called Software as a Service (SaaS) from a hardware point of view. They provide the idea of infinite computing resources available on demand (there is no need to plan their provision in anticipation), the elimination of an up-front commitment by Cloud users (the latter gives the companies possibilities to start small and increase the resources when needed), the ability to pay for the use of computing resources on a short-term basis as needed, i.e. processors by the hour and its storage by the day.

There are other concepts similar to the Cloud computing such as Grid computing. The authors Foster et. al. [33] go through and compare the characteristics of Cloud computing with Grid Computing. Sultan, [32], however, sees things from another perspective, i.e. from a technological point of view, and mentions that the Cloud computing is supported with both virtualization and Grid computing. Via Virtualization it has become possible to use a PC with windows and another operating system such as Linux, or for example to use server virtualization i.e. virtual memory and virtual CPU. In addition it enables single physical resources to behave as multiple logical resources. Grid computing is about the use of several computers to solve a single problem, which requires high degree of processing power. The authors therefore mention that Grid computing can be seen as the technology that provides the establishment of network-distributed parallel processing as well as distributed cluster computing. In addition Cloud computing relies on other technologies such as the Internet and the Web Services as well. These are the technologies that are crucial for the support/hold up of Cloud computing [34].

One of the reasons why Cloud computing has become so popular can be attributed to the fact that it can give scalable IT infrastructure, Quality of Service-assured, as well as, adaptable computing environment [35]. The authors present, as well, a literature review where the authors provide a comprehensive overview on Cloud computing and open federated Cloud computing, as well as, the challenges in Cloud computing. The authors mention that the Federated Cloud computing now attracts more attention, especially the one called the hybrid Cloud, since it provides companies with the possibility to manage their resources in-house, i.e. in a private Cloud and to have other resources externally, i.e., public Cloud. This is convenient for many companies which might have put high investment in their own IT infrastructure, providing the alternative to have sensitive/classified data and information in-house to guarantee its security. Consequently the security and privacy are one of the main challenges in Cloud computing, in concert with scalability. Marston et. al. [30] mentions that the evolution of Cloud computing is one of the major advances in computing, however, in order to be able to achieve its potential there is a need of a clear understanding of different issues involved as well as different perspectives, i.e., the service providers and the users of the technology, i.e. the consumers.

6. DISCUSSION

Integration of data and information is important in decision making process since it greatly simplifies it. Consequently, the Web technologies, especially the Web services, provide new possibilities to integrate data and information, diminishing the importance of the need of a centralised database respectively [13]. The Web technologies, especially the Web 2.0 provide new possibilities for the maintenance engineer, i.e. the possibilities to share knowledge and by doing so provide possibilities to learn and increase the organisational knowledge. This was discussed in Campos and Jantunen [26] where several maintenance tasks are believed to be enhanced through the use of the Web 2.0 technologies, especially the Wikis, when a mobile device is used by the maintenance engineer while he performs his maintenance tasks. For instance, when he needs some specific information at the plant floor, and when a new procedure of the work task is done and inserted into the system, later it can be accessed when needed.

As mentioned before, it is important to understand how companies can benefit from the various ICTs such as the Web technologies, Web 2.0, Cloud computing etc., for this purpose the authors, Marston et. al. [30] provide a roadmap for Information Systems professionals and Information Systems researchers in order to understand and evaluate Cloud computing. They also mention that the benefits of the use of Cloud computing range from lower costs, access to immediate hardware resources, lower ICT barriers to innovation, easier to scale various services, and to acquire applications and services which might not have been possible before. All these benefits of course depend on a specific company and can differ from domain to domain. It would, therefore, be interesting to know how all these factors affect the manufacturing companies practising Condition Based Maintenance (CBM) in their maintenance departments, and how they could benefit from the Cloud computing approach.

In maintenance the use of Cloud computing would lead to the questions on how the Maintenance as a service (MaaS) should be handled. This is because there are many factors to consider when the Cloud computing is used such as the security and scalability. It is important to study the security threats that might appear when Cloud computing is implemented, since in this case sensitive data might not be stored in house and within the same trusted domain. In addition, many of the security approaches that have been evolving during the last 30 years are not valid for the Cloud computing, since the data owner and Cloud computing servers are normally in two different domains. Therefore, various efforts have been made in this direction and emerging system models and security models are being discussed and tested, specific for system with the characteristics of Cloud computing. These include for instance cryptographic methods and data decryption keys [36, 37].

Cloud service providers assert they are able to provide scalability based on the needs of the company and specific department, whenever needed. However, it's important to be aware that the scalability issues differ in Cloud computing, since the system is expected to handle dynamic performance, i.e. changing load regarding performance. There are researches done on the development of benchmark for software system such as the Cloud computing [38], since there is a growing need of tools to measure for instance the load regarding the performance, fault-tolerance and load-balancing etc. of the Cloud computing services. However, condition monitoring is one of the processes in maintenance that produces a huge amount of data that should be processed and analysed, which in turn requires high degree of processing power. For this reason the use of a combination of Grid computing and Cloud computing can be used for this purpose.

Anyway, what could be expected in future is that the Cloud computing services increase in use, in general, since the services that the Cloud computing offers are extremely useful. This can be attributed to the fact that they suggest new ways to handle all the ICT services and provide the company with the latest technology. Moreover, there are cost benefits since the company eliminates some functions that the traditional IT department had. Finally, no matter which ICTs or ICTs services are chosen for the maintenance department, it should be done by taking into account their various characteristics in order to use them optimally and avoid possible disadvantages.

7. CONCLUSIONS

In conclusion, the attempts to standardise different modules of a condition monitoring system and communication and information exchange among them are being made by OSA-CBM. Such systems use artificial intelligence for decision support. The Web technologies, such as Web services, promise several advantages, which lie in the possibility of integrating heterogeneous data and geographically dispersed systems. Moreover, promises the introduction of the Web Services, Semantic Web, Web 2.0 and Social Media Technologies with the core ideas of knowledge management and learning to be used in the e-maintenance applications. The literature findings show that some authors mention several benefits of using the recently emerged approach Cloud computing. It is however important to investigate how

these ICT developments can affect the maintenance departments with their own specific characteristics.

8. REFERENCES

- [1] Liao, S.H. (2005). "Expert system methodologies and applications - a decade review from 1995 to 2004". *Expert Systems with Applications*, 28(1), pp. 93-103.
- [2] Kalogirou, S.A. (2003). "Artificial intelligence for the modeling and control of combustion processes: A review". *Progress in Energy and Combustion Science*, 29(6), pp. 515-566.
- [3] Wang, K. (2003). "Intelligent Condition Monitoring and Diagnosis System A Computational Intelligent Approach". *Frontiers in Artificial Intelligence and Applications*, pp. 93-132, ISBN: 1-56803-312-3.
- [4] Warwick, K., A.O Ekwue and R. Aggarwal, (Eds.). (1997). "Artificial Intelligence Techniques in Power Systems", Power & Energy, Publishing & Inspec, ISBN: 0 85296 897 3.
- [5] Sycara, K.P. (1998). "MultiAgent Systems". *AI Magazine*, 19(2).
- [6] Rao, M., H. Yang and H. Yang (1996). "Integrated distributed intelligent system for incident reporting in DMI pulp mill". *Success and Failures of Knowledge-Based Systems in Real World Applications. Proceedings of the First International Conference*, pp. 169-78.
- [7] Rao M., J. Zhou and H. Yang (1998). "Architecture of integrated distributed intelligent multimedia system for on-line real-time process monitoring". *SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics*, Vol. 2, pp. 1411-16.
- [8] Rao, M., H. Yang and H Yang (1998). "Integrated distributed intelligent system architecture for incidents monitoring and diagnosis". *Computers in Industry*. Vol. 37, pp. 143-51.
- [9] Reichard, K.M., Van Dyke, M. and Maynard, K., (2000). "Application of sensor fusion and signal classification techniques in a distributed machinery condition monitoring system, in Proceedings of the SPIE — The International Society for Optical Engineering, pp. 329–336.
- [10] Thurston, M.G. 2001. "An open standard for Web-based condition-based maintenance systems", 2001 IEEE Autotestcon Proceedings. IEEE Systems Readiness Technology Conference, pp. 401-15
- [11] Muller A, Marquez C & Iung B. (2008). "On the concept of e-maintenance: Review and current research". *Journal of Reliability Engineering and System Safety*, 93(8), pp. 1165–1187.
- [12] Campos, J. & Prakash, O. (2006). "Information and communication technologies in condition monitoring and maintenance", in Dolgui, A., Morel, G and Pereira, C.E. (Eds.) *In-formation Control Problems in Manufacturing*.

Elsevier. Vol. II, pp. 3-8.

- [13] Campos, J. (2009). "Development in the application of ICT in condition monitoring and maintenance", *Journal of Computers in Industry*, Elsevier, 60(1), pp. 1-20.
- [14] Crespo-Márquez A. (2008). "The Maintenance Management Framework: Models and Methods for Complex Systems Maintenance", *Springer Series in Reliability Engineering*.
- [15] Iung, B., Levrat, E., Crespo-Márquez, A. and Erbe, H. (2007). "A first conceptual frame-work for E-maintenance", *IFAC International Conference Cost Effective Automation in Networked Product Development and Manufacturing*.
- [16] Levrat, E., Iung, B. and Crespo Marquez, A.(2008). "E-maintenance: review and conceptual framework". *Production Planning & Control*, Vol. 19, No. 4, pp. 1-22.
- [17] Iung, B. (2006). "Research and development on e-maintenance: Survey of some recent ad-vances and directions". *International Conference on Maintenance Management*, pp. 95-100.
- [18] Holmberg K., Adgar A., Arnaiz A., Jantunen E., Mascolo J. & Mekid S. (2010). "E-maintenance", Springer, ISBN 978-1-84996-204-9
- [19] Karim R. (2008). "A Service-Oriented Approach to eMaintenance of Complex Technical Systems". *Doctoral Thesis*, Luleå University of Technology, 2008:58, ISSN: 1402-1544
- [20] Campos Jeria, J. (2008). "ICT tools for e-maintenance". *Doctoral Thesis Växjö University*. 2008:151, ISSN: 1404-4307.
- [21] Hendler J., Berners-Lee T and Miller E. (2002) "Integrating Applications on the Semantic Web", *Journal of the Institute of Electrical Engineers of Japan*, 122(10), pp. 676-680.
- [22] Win T. (2006). "Web 2.0: is it really different?". *Magazine Networker*, 10(2), pp. 15-17.
- [23] Aharony, N. (2009). "The influence of LIS students personality characteristics on their perceptions towards Web 2.0 use". *Journal of librarianship and information science*, 41(4), pp. 227-242.
- [24] Boateng, R., Mbarika, V & Thomas, C. (2010). "When Web 2.0 becomes an organizational learning tool: evaluating Web 2.0 tools". *Development and learning in organizations*, 24(3), pp. 17-20.
- [25] McGee, J.B & Begg, M. (2008). "What medical educators need to know about Web 2.0". *Medical teacher*, 30(2), pp. 164-169.
- [26] Campos, J and Jantunen, E (2011). "The Semantic Web and Web 2.0 for e-maintenance applications". *Proceedings of the 24th International Congress on Condition Monitoring and Diagnostics Engineering Management, COMADEM 2011*, pp. 463-473.
- [27] Craig, E.M. (2008). "Changing paradigms: managed learning environments and Web 2.0". *Campus-Wide Information Systems*, 24(3), pp. 152-61.
- [28] Xu, X. (2012). "From cloud computing to cloud manufacturing". *Robotics and Computer-Integrated Manufacturing* Vol. 28, pp. 75-86.
- [29] Wang, L & Laszewski, G. (2008). "Scientific Cloud Computing: Early Definition and Experience". *Service Oriented Cyber infrastructure Lab, Rochester Institute of Technology*.
- [30] Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., & Ghalsasi, A. (2011). "Cloud computing— The business perspective". *Decision Support Systems* Vol. 51, pp. 176-189.
- [31] Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R.H., Konwinski, A., Lee, G., Patterson, D.A., Rabkin, A., Stoica, I. & Zaharia, M. (2009). "Above the Clouds: A Berkeley View of Cloud Computing". *Tech. Rep. UCB/ECS-2009-28, EECS De*
- [32] Sultan, N. (2011). "Cloud computing for education: A new dawn?", *International Journal of Information Management* Vol. 30, pp. 109-116.
- [33] Foster, I., Zhao, Y., Raicu, I., & Lu, S. (2008). "Cloud Computing and Grid Computing 360-Degree Compared". *Proc. IEEE Grid Computing Environments Workshop*, pp. 1-10.
- [34] Carr, N. (2008). "The Big Switch: Rewiring the World, from Edison to Google". *New York & London: W.W. Norton*.
- [35] Yang, X., Nasser, B., SurrIDGE, M., & Middleton, S. (2012). "A business-oriented Cloud federation model for real-time Applications". *Future Generation Computer Systems*, Vol. 28, pp. 1158-1167
- [36] Yu, S, Wang, C., Ren, K. & Lou, W. (2010). "Achieving secure, scalable, and fine-grained data access control in cloud computing". In *Proceedings of IEEE INFOCOM 2010*, pp. 15-19.
- [37] Tysowski, P & Hasan, M. A. (2011). "Towards secure communication for highly scalable mobile applications in cloud computing systems". *Centre for Applied Cryptographic Research University of Waterloo, Tech Rep*, Vol. 33.
- [38] Binnig, C., Kossmann D., Kraska T & Loesing S. (2009). "How is the Weather tomorrow? Towards a Benchmark for the Cloud". In *Proceedings of the 2nd International Workshop on Testing Database Systems (DBTest '09)*.

Case-Based Reasoning Supports Fault Diagnosis Using Sensor Information

Ning Xiong
School of Innovation, Design, and
Engineering
Mälardalen University
Västerås, Sweden
+46 21 151716
ning.xiong@mdh.se

Tomas Olsson
Industrial Applications and
Methods (IAM)
Swedish Institute of Computer Science
SICS, Kista, Sweden
+46 70 783 1547
tol@sics.se

Peter Funk
School of Innovation, Design, and
Engineering
Mälardalen University
Västerås, Sweden
+46 21 103153
peter.funk@mdh.se

ABSTRACT

Fault diagnosis and prognosis of industrial equipment become increasingly important for improving the quality of manufacturing and reducing the cost for product testing. This paper advocates that computer-based diagnosis systems can be built based on sensor information and by using case-based reasoning methodology. The intelligent signal analysis methods are outlined in this context. We then explain how case-based reasoning can be applied to support diagnosis tasks and four application examples are given as illustration. Further, discussions are made on how CBR systems can be integrated with machine learning techniques to enhance its performance in practical scenarios.

Keywords

Fault Diagnostics, Case-based Reasoning, Sensors, Signal Processing, Feature Extraction, Crack Detection, Process Monitoring, Artificial Intelligence, Knowledge Discovery, Case Retrieval

1. INTRODUCTION

A fault is an abnormal state of a machine or a system such as dysfunction or malfunction of a part, an assembly, or the whole system. As machines become larger and more complex with industrial development, the costs and technical know-how required for system maintenance increases substantially. Fast and precise identification of faults and problems in equipment makes a crucial contribution to the enhancement of reliability in manufacturing and efficiency in product testing.

Sensor data provides rich and objective information for fault diagnosis in monitoring and maintenance. An experienced technician has skills to find symptoms for detecting an existing fault or predicting a failure in near future. However a junior technician may fail to do this due to the lack of knowledge and experiences. Artificial intelligence techniques offer valuable tools for building intelligent, computer-based systems for monitoring and diagnosis that are based on human knowledge and past experiences [12]. Such computer-based systems can offer decision support in practical applications by offering a second opinion.

Regarding pattern classifier for monitoring and diagnosis, a number of methods might be considered. Expert systems were developed in support of gathering, representing and utilizing human expert knowledge for problem solving but they suffer from

the knowledge acquisition bottleneck. Regression functions distinguish objects by defining linear boundaries between classes using a moderate number of attributes as function variables. For problems with nonlinear boundaries artificial neural networks would be a candidate approach because they are capable of realizing arbitrary nonlinear mappings between input and output units. Nevertheless the functions of neural networks are like a black box, they hardly provide any reasons and arguments for decisions recommended by them. Comparatively case-based reasoning (CBR) [1] is more transparent by making decisions according to similar cases retrieved such that human users are given reference information to understand, verify, and occasionally also modify the suggested results. The explanatory issue is quite important in industrial applications where AI systems serve as decision support and every decision made has to be well justified before taking into effect. This explains why this paper focuses on the use of CBR to classify time series signals for fault diagnosis in industrial scenarios.

The remaining of the paper is organized as follows. Section 2 gives an overview of the methods of signal analysis as a prior step for CBR. The CBR approach to fault diagnosis is introduced in section 3, which is followed by several application examples in section 4. In section 5 we discuss some ways to further improve CBR performance in diagnosing tasks. Finally section 6 gives concluding remarks.

2. ANALYSIS OF SENSOR SIGNALS PRIOR TO DIAGNOSIS

Abnormality of industrial machines can be reflected by some key states during their operation. Using sensor technology it is possible to detect and measure the values of these system states and their profiles. We can then process and analyse the collected sensor recordings in order to find out hidden symptoms. Reasoning can be made based on detected symptoms to estimate the class of fault associated with the machine or make prediction about what potential problem is likely to occur in a near future. A general paradigm for signal-based diagnosis is illustrated in Fig. 1, which includes signal filtering, feature extraction, and pattern classifier as its important components.

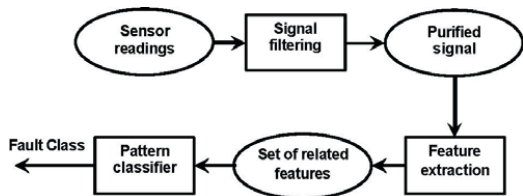


Figure 1. Fault diagnosis based upon sensor signals

Signal filtering is used to purify original sensor readings by removing the noises contained in the signals such that more reliable diagnosis results will be warranted. Usually there are two kinds of noises involved in the perceived signals; one is measurement noise due to intrinsic imprecision of sensors and the other is external noise caused by disturbance from surroundings and which is added to the sensor data received. Signal recovery from external background noise has been well dealt with by applying signal processing methods like wavelet analysis and time domain averaging [21][22]. Reduction of measurement errors is outside the scope of this paper, but interested readers can refer to sensor fusion systems in which Bayesian based filtering approaches such as Kalman filtering [5] and particle filtering [15] are worthy of being used to acquire more accurate estimates of the states of the underlying machines or processes.

Feature extraction aims to identify characteristics of sensor signals as useful symptoms for further analysis. This stage is crucial in many practical domains in industry where the process in consideration is dynamic such that measured states generally change with the time. This means that it is not possible to depict sensor observations with static single values. Instead we need to identify a group of features to characterize a time series signal. The set of extracted features is desired to have a moderate number to facilitate efficient analysis and reasoning. On the other hand, features extracted also ought to be adequate to accommodate temporal information or transitional patterns of signals to be analysed.

Conventionally features extracted from signals fall into two categories, namely statistical features and frequency-based features. Statistical features are extracted from the profile of signal values with respect to calculated statistics as overall generalization. Typical features of this kind can be peak value, start time, overshoot, rising time, mean value, integral, standard deviation, etc. In practice what features to use for signal representation is commonly ad-hoc and domain dependent. An example of using statistical features for case-based circuit diagnosis was illustrated in [27]. However one converts dynamic data streams into static values when making statistical features, which may lead to the loss of information about temporal relation between data.

Frequency-based features characterize sensor signals by groups of quantities related to a diversity of frequencies. As numerous signal transforms are available to yield frequency spectra, we seem to have more solid basis for extracting features based on frequency than for deriving features based on statistics. The two most common signal transform methods are Fourier Transform and Wavelet Analysis **Error! Reference source not found.**

However traditional feature extraction methods may have some drawbacks such as large number of features as well as the risk of losing temporal relationship existing in the original signals.

Symbolic approximation was proposed in [13] as a more compact and meaningful way to characterize the dynamic property of complex, longitudinal series of measurements. The solution is to convert the sampling-point based representation of the time series into an interval-based representation. An interval consists of a set of consecutive sampling points and thus encompasses multiple sampling periods in the time dimension. Then data within an interval have to be generalized and aggregated into one symbolic value; the symbolization is conducted via discretization of the range for possible values of the signal. By doing this, the primary time series is transformed into a symbolic series associated with intervals. Symbolic approximation of primary numerical time series signals would bring the following benefits:

- Symbolic series are shorter in length and more intensive in information content, while much of the important temporal information is still retained.
- Symbolic series facilitate higher computational efficiency and require less computational resource and memory space.
- Symbolic data are more robust and less sensitive to measurement noises.
- Symbolic data are easier for human understanding and inspection.

Temporal abstraction [30] from Artificial Intelligence provides a good technique to implement symbolic approximation, i.e., to transform time-stamped numerical signals into interval-based symbolic representations. It works by aggregating adjacent events exhibiting a common behaviour over time into a generalized concept. The ontology for basic temporal abstraction includes state abstraction and trend abstraction. The former focuses on the measured values themselves to extract intervals associated to qualitative concepts such as low, normal, and high, while the latter considers differences between two neighbouring records to detect specific patterns like increase, decrease, and stationarity in the series.

3. CASE-BASED REASONING FOR DIAGNOSIS

Case-based reasoning (CBR) offers an effective means to implement pattern classifier. Motivated by the doctrine that similar situations lead to similar outcomes, CBR fits well to classify the current new sensor signals based on experiences of past categorizations [24]. The main strength of CBR lies in the fact that it enables directly reusing concrete examples in history and consequently eases the knowledge acquisition bottleneck. It also creates the opportunity of learning from experiences but skipping the step of data training such that the over-fitting problem no longer exists.

We perform CBR to make classification of faults using known cases in the case library as shown in Fig. 2. It starts with similarity matching to compare the query problem against cases stored in the case base and a set of similar cases are thereby retrieved. Then, in the next stage, the solutions (classes) of

retrieved similar cases are combined in terms of their similarity values in order to estimate the fault class of the new problem.

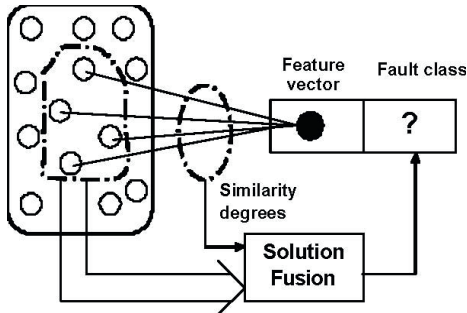


Figure 2. Fault diagnosis based upon sensor signals

Generally the working cycle for a CBR system consists of the following four steps: retrieve, reuse, revise and retain, as shown in Fig. 3. The first step is to retrieve a set of similar cases given a new problem. It is followed by the reuse step, which is tasked to reuse the information from the retrieved cases to suggest a new solution to the query problem. Usually the retrieved cases are not identical to the query problem; we need to perform adaptation or fusion of the retrieved solutions to fit the new situation. The third step is revise, in which the suggested solution will be verified for its correctness. Unsuitable solutions will be revised and re-verified in this step. Finally, in the retain step, a confirmed solution will be treated as a learned new case and stored into the case library for future usage.

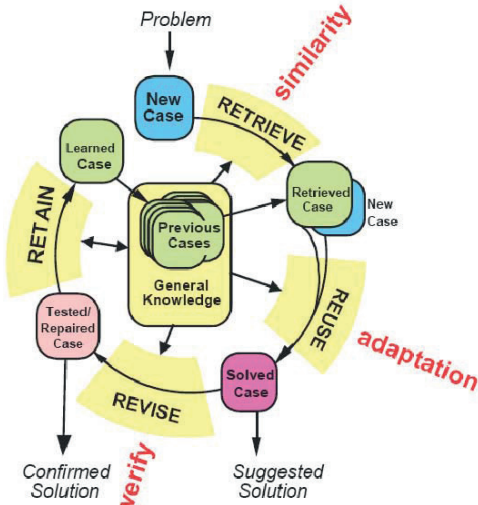


Figure 3. The CBR Cycle

Generally the working cycle for a CBR system consists of the following four steps: retrieve, reuse, revise and retain, as shown in Fig. 3. The first step is to retrieve a set of similar cases given a new problem. It is followed by the reuse step, which is tasked to reuse the information from the retrieved cases to suggest a new solution to the query problem. Usually the retrieved cases are not identical to the query problem; we need to perform adaptation or fusion of the retrieved solutions to fit the new situation. The third step is revise, in which the suggested solution will be verified for its correctness. Unsuitable solutions will be revised and re-verified in this step. Finally, in the retain step, a confirmed solution will be treated as a learned new case and stored into the case library for future usage.

4. INDUSTRIAL APPLICATION EXAMPLES USING CBR

In this section we briefly highlight several projects of applying CBR techniques for process monitoring, diagnosis and quality control, which are being performed or have been completed by the Intelligent Systems Group at Mälardalen University, Sweden.

4.1 Crack detection in welding process

In the project about crack detection, recordings of several welding processes were done mainly focusing on the cool down time in the near seconds after a finished welding process [19]. The aim is to determine the signature that is generated by an emerging crack during the welding process.

Some results from the case-based crack detection project are the following:

- Ultra sound sensor(s) were selected as a suitable recording equipment
- A set of initial recordings from different phases of welding processes was collected and then the data from 2 normal and 10 welds with cracks were analysed.
- Suitable features to classify the condition of welding processes were selected
- Methods and algorithms for classification of welding processes were developed. The result is that 100% of cracked welds were correctly classified and only one normal weld was incorrectly classified as cracked.

4.2 Process monitoring and diagnosis of milling machines

The milling machine project focused on sensorless, nonintrusive monitoring of a milling process. The sensor data used were sound measurements recorded during the milling process. Adequate features were identified from the sound signals and they were used as inputs to represent cases in the case-based diagnostic system.

Cases were created from recorded signals that contain typical patterns of cutting conditions and cutting stages of the milling machine. We then utilized these cases in a CBR procedure to estimate the status of the machine given new sensor measurements. The condition of the milling process can be classified into a category that is associated with a fault.

4.3 Geometric production adjustments

The on-going project is in collaboration with Volvo Car (in Sweden) with the objective of finding proper adjustment actions to produce precise geometric shape of car bodies in terms of specifications. A case-based tool has been developed for supporting the adjustments of a production line to prevent producing parts of the body drifting towards unacceptable dimensions (called defect parts) [25]. Measurements, adjustments and their outcome of defective parts are connected inside a case. A case library of such cases has been created and made available for real-time decision support. The reasoning with cases in the case library provides a mapping from measurements to suitable actions to correct the production back towards to the desired state.

4.4 Robot fault diagnosis

A CBR approach for diagnosing faulty robot gearboxes was developed in [14] and [26]. In the project both Discrete Wavelet Transform and Discrete and Fast Fourier Transforms were used in order to extract signal features. Then, a new signal a query was classified using a CBR approach based on cases of known faulty and normal gearboxes in the case library.

5. FURTHER ENHANCING CBR PERFORMANCE

Although CBR is a powerful methodology to facilitate problem solving by resorting to previous similar experiences, the functionality of a CBR system may not always be optimal in practical applications. This is owing to the following two reasons. First, the cases available can be limited and hence they cannot cover the whole problem space. Secondly, CBR is usually performed in areas with poor domain theory and therefore there is no precise knowledge for guiding the utilization of individual cases in a case-based inference procedure. In the remaining of this section we shall discuss these two possible limitations of CBR could be compensated by integration with model-based reasoning and knowledge discovery.

5.1 Integration with Model-Based Reasoning

It is frequently the circumstance that neither the cases at hand nor the available knowledge model can achieve a complete coverage of the problem space. Yet both can complement each other to solve problems in broader scope. Hence it is a natural practice to build a hybrid system with integration of both case-based and model-based reasoning in a cooperative manner to be able to identify more faults and problems.

In the INRECA project [3][4], the knowledge model of causal and decision trees were combined with CBR for applications in robot fault diagnosis. A CBR system combined with expert system was presented in [18] to detect defects in images from ultrasonic rail-inspections. The images were first classified by a set of expert rules. If the classification failed, then the CBR system was used to solve the classification task. The system in [11] hybridized CBR with model-based reasoning in two scenarios: diagnosis for robots and diagnosis of a nuclear ventilation system. The method of CBR was used in the system to help users finding alternative solutions as a complement to the model-based suggestions as well as to enable the system to learn from experience.

5.2 Discovery of CBR Knowledge

In order to conduct CBR optimally, intensive knowledge is required to guide the way of dealing with cases. For instance, retrieve knowledge is needed as criteria to judge which cases are similar and useful for solving a new problem. For case indexing, we need the knowledge and information for what attributes to adopt to concisely characterize the nature of a case. All such knowledge for CBR is domain related. In the following we shall discuss how to acquire the knowledge for case representation and case retrieval respectively via data mining and knowledge discovery.

5.2.1 Discovery of knowledge for case representation

We prefer a concise case index to keep a low input dimension for the case-based classifier. However, traditional methods of feature extraction usually produce a large number of features from the original signals. Some features may be irrelevant, redundant, or contaminated by heavy noise. With feature selection we intend to derive the knowledge about which features are important based on the given data set. Only important features are selected as inputs for building the case index. Selection of important features leads to improved CBR performance, lower input dimensionality, as well as reduced computational costs.

Existing approaches to feature selection can be divided into two categories, called filter and wrapper respectively. Filter approaches [28], [10], try to assess feature goodness as an intrinsic property independent of the modelling algorithm. Usually they attempt to detect possible dependency between a pair of variables based on the given data. But individual features are evaluated in isolation of each other without considering the influence of others. In contrast, wrapper approaches [20], [23], [16] use a modelling algorithm to evaluate feature subsets in terms of the modelling accuracy. Hence they may yield higher classification accuracy than filtering approaches.

It is interesting to mention that feature selection and CBR can support each other. In [33] and [35] CBR was used as a criterion to evaluate the goodness of a feature subset. This proposed approach seems similar to wrapper but it reduces the complexity of wrapper by requiring no modelling procedure given a feature subset. Later, the paper [32] demonstrated that feature selection could be conducted to enhance CBR systems to reach high classification accuracy.

The other challenging issue for case representation lies with symbolic sequential data (after symbolic approximation of original signals), where the existing numerical approaches for signal processing are not applicable. This issue was tackled in [13] by focusing on transitions of states in time series. It is believed that in many real-world situations, key transitions of states are more important than symbolic values themselves in detecting possible faults. Such transitions of states are referred to as key sequences. It was suggested in [13] that the knowledge about key sequences being discovered based on a database of classified time series.

Once the key sequences are identified, they are utilized as reference to capture important contents in a time series of query. The symbolic series is checked thoroughly to detect any occurrences of key sequences in it. Then the information about whether a key sequence has occurred and with how many times is made use of in building a numerical case index. This case index is

concise since it only considers appearances of key sequences while ignoring other trivial randomness. Various ways to construct such case indexes have been addressed in [13].

5.2.2 Discovery of knowledge for case retrieval

Similarity assessment plays an important role in the retrieval step of a CBR cycle. As the fundamental principle for CBR is that similar problems have similar solutions, we expect to rely on the similarity model to retrieve the truly useful cases for solving a new problem. It was pointed out in [29] that a competent similarity model should function as a knowledge container by encoding domain knowledge.

So far the most common way of similarity modelling has been focused on feature weighting [31]. Features are assigned with different weights in accordance with their importance, and the global similarity metric is defined as a weighted sum of the local matching values in single attributes. Different approaches of interest have been proposed for identifying such weights automatically. Incremental learning attempts to modify feature weights according to success/failure feedback of retrieval results [6]. The probability of ranking principle was utilized in [8] for the assignment of weight values to features. Case-ranking information was utilized in [7] and [9] for weight adaptation towards similarity degrees of retrieved cases consistent with a desired order. Accuracy improvement represents another way for adapting the set of weights as discussed in [17] and [2]. Nevertheless, no matter how the values of weights are derived, the capability of these similarity-learning methods is inherently constrained by weighted combination of the local matching degrees. This limitation in the structure of similarity makes it hard to represent more general knowledge and criteria for case assessment and retrieval.

A new similarity model without feature weighting was proposed in [34] as an effort to seek more powerful representation of knowledge for case retrieval. The idea was to encode the information about feature importance into local compatibility measures such that feature weighting is no longer needed. Later, it was also analysed and demonstrated that the parameters of such compatibility measures can be learned from the case library in favour of coherent matching, i.e. to maximize the supportive evidence while minimize the amount of inconsistency derived from pairwise matching of cases from the case base.

6. CONCLUSIONS

Applying AI techniques for signal-based fault diagnosis presents an emerging trend in maintenance research. Interesting decision support systems can be built by using AI techniques to classify collected signal patterns. Case-based reasoning and supervised learning are two powerful AI techniques to classify sensor data in terms of faults and also improve system performance when experiences accumulate. The advantages of case-based reasoning include simplicity, easy understanding, good explanation, as well as the ability to survive with a small number of cases. On the other hand, supervised learning represents well-established and sound approaches to learning and generalization from a number of training examples. They can be trained to approximate training examples with satisfying accuracy. However, a weakness with supervised learning methods is that they require sufficient examples for training to avoid the risk of overfitting, hence they

are not applicable in diagnosis tasks where the amount of gained experiences is rather limited.

ACKNOWLEDGEMENTS

This work has been funded by the CREATE project (ITEA2) and by the Swedish Foundation for Strategic Research (SSF, ProViking).

REFERENCES

- [1] A. Aamodt and E. Plaza, Case-based reasoning: foundational issues, methodological variations and systems approaches, *AI Communications*, vol. 7, pp. 39-59, 1994.
- [2] H. Ahn, K. Kim, and I. Han, Global optimization of feature weights and the number of neighbors that combine in a case-based reasoning system, *Expert Systems*, vol. 23, pp. 290-301, 2006.
- [3] K. Altho et al., Case-based reasoning for decision support and diagnostic problem solving, In: *Proceedings of the Third German Conference on Knowledge-Based Systems (XPS-95)*, 1995.
- [4] E. Auriol et al., Integrating case-based reasoning and hypermedia documentation: an application for the diagnosis of a welding robot at Odense steel shipyard, *Engineering Applications of Artificial Intelligence*, vol. 12, pp. 691-703, 1999.
- [5] R. Bar-Shalom, Y. and X. Li, *Estimation and Tracking: Principles, Techniques, and Software*, Artech House, Boston, 1993.
- [6] A. Bonzano, P. Cunningham, and B. Smith, Using introspective learning to improve retrieval in CBR: A case study in air traffic control, In: *Proc. 2nd International Conference on Case-based Reasoning*, Providence RI, USA, 1997, pp. 291-302.
- [7] K. Branting, Acquiring customer preferences from return-set selections, In: *Proc. 4th International Conference on Case-Based Reasoning*, 2001, pp. 59-73.
- [8] N. Cercone, A. An, and C. Chan, Rule-induction and case-based reasoning: Hybrid architectures appear advantageous, *IEEE Trans. Knowledge and Data Engineering*, vol. 11, pp. 166-174, 1999.
- [9] L. Coyle, and P. Cunningham, Improving recommendation ranking by learning personal feature weights, In: *Proc. 7th European Conference on Case-Based Reasoning*, 2004, pp. 560-572.
- [10] M. R. Emami and I. B. Türksen, Development of a systematic methodology of fuzzy logic modelling, *IEEE Trans. Fuzzy Systems*, vol. 6, pp. 346-361, 1998.
- [11] M. Feret and J. Glasgow, Combining model-based and case-based reasoning for the diagnosis of complex devices, *Applied Intelligence*, vol. 7, pp. 57-78, 1997.
- [12] P. Funk, and N. Xiong, Why we need to move to intelligent and experience based monitoring and diagnostic systems, In: *Proceedings of 23th International Conference on Condition Monitoring and Diagnostic Engineering Management*, 2010, pp. 111-115.

- [13] P. Funk, and N. Xiong, Extracting knowledge from sensor signals for case-based reasoning with longitudinal time series data, In: *Case-Based Reasoning in Signals and Images*, P. Perner (Eds.), Springer, 2008, pp. 247-284, 2008.
- [14] P. Funk, E. Olsson, M. Bengtsson, and N. Xiong, Case-based experience reuse and agents for efficient health monitoring, prevention and corrective actions, In: *Proceedings of 19th International Congress on Condition Monitoring and Diagnostic Engineering Management*, Luleå, Sweden, 2006, pp. 445-453.
- [15] N. Gordon, A. Marrs and D. Salmond, Sequential analysis of nonlinear dynamic systems using particles and mixtures, In: *Nonlinear and Nonstationary Signal Processing*, W. Fitzgerald, R. Smith, A. Walden, and P. Young, ed., Chapter 2, Cambridge University Press.
- [16] C. N. Hsu, H. Huang, and S. Dietrich, The ANNIGMA-wrapper approach to fast feature selection for neural nets, *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 32, no. 2, pp. 207-212, 2004.
- [17] J. Jarmulak, S. Craw, and R. Rowe, Genetic algorithms to optimize CBR retrieval, In: *Proc. European Workshop on Case-Based Reasoning (EWCBR 2000)*, 2000, pp. 136-147.
- [18] J. Jarmulak, E. Kerckhoffs and P. van't Veen, Case-based reasoning in an ultrasonic rail-inspection system, In: *Case-Based Reasoning Research and Development*, Lecture Notes in Computer Science, ed. Leake D and Plaza E (Springer Berlin / Heidelberg), pp. 43-52, 1997.
- [19] L. Karlsson, Crack Detection in welding process using acoustic emission, Master's thesis, School of Innovation, Design and Engineering, Mälardalen University, 2010.
- [20] R. Kohavi and G. H. John, Wrapper for feature subset selection, *Artificial Intelligence*, vol. 97, no. 1/2, pp. 273-324, 1997.
- [21] S. K. Lee and P. R. Whit, The Enhancement of impulse noise and vibration signals for fault detection in rotating and reciprocating machinery, *Journal of Sound and Vibration*, vol. 217, pp.485-505, 1998.
- [22] J. Lin, Feature extraction of machine sound using wavelet and its application in fault diagnosis, *NDT&E International*, vol. 34, pp. 25-30, 2001.
- [23] K. Z. Mao, Feature subset selection for support vector machines through discriminative function pruning analysis, *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 1, pp. 60-67, 2004.
- [24] T. Olsson, P. Funk, Case-based reasoning combined with statistics for diagnostics and prognosis, In: *Proc. 25th International Workshop on Condition Monitoring and Diagnostic Engineering*, 2012.
- [25] E. Olsson, P. Funk and A. Andersson, Case-based reasoning applied to geometric production measurements, In: *Proc. 1st International Workshop and Congress on eMaintenance*, June 22-24, 2010, pp. 25-28.
- [26] E. Olsson, P. Funk, and N. Xiong, Fault diagnosis in industry using case-based reasoning, *Journal of Intelligent and Fuzzy Systems*, vol. 15, pp. 41-46, 2004.
- [27] C. Pous, J. Colomer, and J. Melendez, Extending a fault dictionary towards a case based reasoning system for linear electronic analog circuits diagnosis, In: *Proceedings of the 7th European Conference on Case-Based Reasoning*, Madrid, 2004, pp. 748-762.
- [28] K. Praczyk, H. Kiendl, and T. Slawinski, Finding relevant process characteristics with a method for data-based complexity reduction, In: *Lecture Notes in Computer Science 1625*, Springer Verlag, 1999, pp. 548-555.
- [29] M.M. Richter, The knowledge contained in similarity measures, invited talk at the International Conference on Case-Based Reasoning, 1995.
- [30] Y. Shahar, A framework for knowledge-based temporal abstractions, *Artificial Intelligence*, vol. 90, pp. 79-133, 1997.
- [31] D. Wetschereck, and D. Aha, Weighting features, In: *Proc. 1st Int. Conf. Case-based Reasoning*, 1995, pp. 347-358.
- [32] N. Xiong, and P. Funk, Combined feature selection and similarity modeling in case-based reasoning using hierarchical memetic algorithm, In: *Proceedings of the IEEE World Congress on Computational Intelligence*, 2010, pp. 1537-1542.
- [33] N. Xiong, and P. Funk, Construction of fuzzy knowledge bases incorporating feature selection, *Soft Computing*, vol. 10, pp. 796 – 804, 2006.
- [34] N. Xiong, and P. Funk, Building similarity metrics reflecting utility in case-based reasoning, *Intelligent & Fuzzy Systems*, vol. 17, pp. 407-416, 2006.
- [35] N. Xiong, A hybrid approach to input selection for complex processes, *IEEE Trans. Systems, Man, and Cybernetics, Part A*, vol. 32, pp. 532-536, 2002.

Framework for Design and Performance Evaluation of Effective Condition Monitoring and Maintenance Management Systems

Rezsa Farahani
University of Alberta
4-9 Mechanical Engineering
University of Alberta
Edmonton, AB, T6G2G8
+1(780) 492-9494
rfarahani@ualberta.ca

Michael G. Lipsett
University of Alberta
4-9 Mechanical Engineering
University of Alberta
Edmonton, AB, T6G2G8
+1(780) 492-9494
michael.lipsett@ualberta.ca

ABSTRACT

A framework has been developed that extends current information system implementation frameworks for performance-based, fit-for-purpose, and holistic implementation of condition monitoring and maintenance management systems. The framework aims to guide design and implementation of effective systems, to be used as a tool for concretely demonstrating the value of system features, and to ultimately assess system performance after implementation for continuous improvement. An effective system from the framework's point of view is performance-based, i.e. benefits due to the system are balanced with associated costs; is fit-for-purpose, i.e. is designed with respect to organizational business strategy, objectives, and constraints; and is holistic, i.e. addresses various performance criteria, e.g. financial, operational excellence, safety and environmental risk minimization, handling uncertainty, system intelligence and automation.

Keywords

EMaintenance, Maintenance management, enterprise system, performance measurement, information systems, reliability engineering, CMMS, CMMIS, condition based maintenance, condition monitoring, asset information management system, CBM, performance based, holistic.

1. INTRODUCTION

A framework has been developed that extends current information system implementation frameworks for effective implementation of condition monitoring and maintenance management information systems (CMMIS), also known as EMaintenance systems.

These systems have significantly enhanced availability and reliability of equipment by providing relevant and timely information on asset status and health via condition and process monitoring systems, and triggering appropriate responses based on suitable information via control and maintenance systems.

However, an unsuitable system may not be effective in providing the intended level of performance. Furthermore, it is often challenging to determine the value of various system functionalities for system design decision making, and difficult to design performance metrics for system performance evaluation. The framework aims to guide design and implementation of effective CMMIS, to be used as a tool for concretely

demonstrating the value of system features, and to ultimately assess system performance after implementation for continuous improvement.

An effective system, from the framework's point of view, is performance-based, i.e. benefits due to the system are balanced with associated costs; is fit-for-purpose, i.e. is designed with respect to organizational business strategy, objectives, and constraints; and is holistic, i.e. addresses various performance criteria, e.g. financial, operational excellence, safety and environmental risk minimization, handling uncertainty, and intelligence and automation.

2. THE FRAMEWORK

The framework has a strong focus on the information system aspect of the CMMIS. The performance of a CMMIS must be considered as part of its underlying industrial information system (IIS) and with respect to the business context of the organization, i.e. its enterprise system. The influence of IIS on maintenance management is profound considering the central information management role that it plays in many organizations, and which as a result, affects all aspects of organizational performance.

Maintenance work requires working with various sources of information and large amounts of data, communicating with many stakeholders such as customers, vendors, coworkers, and making complex decisions. An effective maintenance system involves coordination of information, processes, and decisions among several different levels of operation, i.e. equipment, field, and the business (organization). Hence modern maintenance management methods possess the rich set of characteristics that require the full potential of an IIS to perform effectively and efficiently. For this reason, the proposed framework puts the IIS in a central position for CMMIS implementation. Any CMMIS implementation framework should consider the IIS capabilities that can affect decision making, learning, knowledge management, collaboration, analysis and intelligence, risk awareness and alerting, visualization, mobility and de-centralized operations, flexibility, interoperability, etc.

The current practice of equipping an industrial organization with IIS typically involves a requirements gathering phase (specification) followed by design, development and implementation phases based on some assessed requirements [1].

This research aims to extend and complement the existing practice by developing a holistic framework for performance-based design, development, and evaluation of CMMIS systems as a continuous improvement process (Figure 1). The framework is also intended to provide a rigorous, systematic assessment process for implementation of a CMMIS.

The developed framework is composed of two pre-design and development activities (outside of the grey area in Figure 1) in addition to the typical specification, design, development and evaluation activities:

1. *Needs assessment*: provides a systematic means in defining the current state of the industrial application (the organization), the requirements for implementing the intended CMMIS, the resources and methods for design and development, and the system performance measurement criteria; and
2. *Performance criteria development*: activities related to designing performance evaluation methods, metrics, or key performance indicators for pre-evaluating potential system performance and gauging performance after implementation.

To achieve the framework objectives, the needs assessment activities work towards developing a fit-for-purpose system. Diversity of the performance criteria enables a holistic system. And the accommodation of value trade-off and performance metrics development and measurement by the framework steer the design towards a performance-based system.

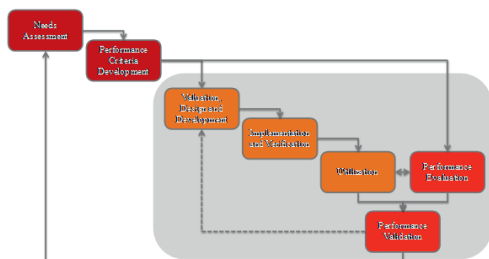


Figure 1. The performance-based system implementation framework for CMMIS: a continuous improvement process.

2.1 Needs Assessment

The needs assessment activities are intended to provide a systematic assessment process for establishing system requirements, its expected performance, and the state of the user organization for implementing the desired performance-based CMMIS. These systematic activities would facilitate practice consistency and the elimination of any arbitrariness in the client needs assessment activities. It would provide a reliable assessment process with clear orientation towards an organization's objectives, and with the purpose of clarifying the system value and the gaps that could be filled by system implementation for all stakeholders. It would also increase the likelihood that all requirements that are critical to the organization's objectives would be addressed by the CMMIS without costly future modifications and re-work. The framework's needs assessment activity involves five assessment topics:

1. *Existing system assessment*: for understanding the organization's business model, operations, information system, and information flow and usage;
2. *Requirements assessment*: for defining objectives that are expected from the system, the business strategy to be respected and enabled by the system, and the design and implementation constraints;
3. *Resource assessment*: for identifying the resources available and required for system design, development and implementation, which primarily includes financial capital, human capital, information, and technology;
4. *Maintenance Management Method assessment*: for assessing processes, instructions, procedures, and logic for achieving maintenance and asset management objectives and accomplishing certain tasks within a specific application. This includes business systems, management systems, engineering or specific function systems, and industrial process systems required for the maintenance management and asset management practice; and
5. *Performance criteria selection*: for selecting performance criteria, which includes the criteria for requirements assessment, related to maintenance management, and other criteria (colors orange, brown, and red in Figure 2, under Performance Criteria Selection, respectively) critical to effective performance of a fit-for-purpose IIS, which have been deemed important or of high priority for the organization (but not explicitly specified as part of requirement assessment and specific domain criteria).

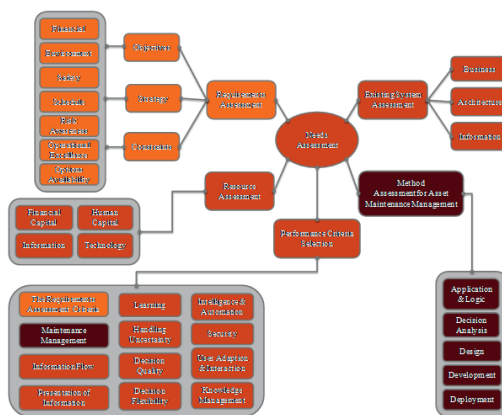


Figure 2. Activity classes (aspects) of the framework's needs assessment component.

2.1.1 Existing System Assessment

One aspect of existing system assessment within the framework is to clarify the context of the business for CMMIS implementation, i.e. its industrial processes and operations, the relevant business and organizational dynamics, relations, and operations including the overall values and objectives of the business; The immediate, short term, and long term competitive strategies; The business market, industry, competitors and stakeholders; Organizational factors such as size, structure, culture, and location; Related to

maintenance: maintenance operations and practices, maintenance costs, general financial indicators such as production, revenue, costs, and profits, and general organizational factors relevant to maintenance; Existing IIS architecture, e.g. whether it's a: silo system, fully integrated service oriented architecture (SOA) system, existence of web-applications, communication network, etc..

In assessing the existing system, the existing information dynamics and flow would also be determined. This is critical in defining system architecture and information dynamics for a new CMMIS, since the new system must effectively and efficiently use the information that already exists in the system, which generally is the foundation of the organization's knowledge and competitive advantage.

2.1.2 Requirements Assessment

System requirements assessment is performed to define the objectives which the organization wishes the system to achieve, to determine constraints and limitations for the system, and the business strategy that would guide and affect these objectives and constraints. Strategy determines the priority and the importance of the objectives and constraints for the organization.

The main criterion for setting objectives and constraints for a new system implementation has traditionally been financial. In worst cases, this means objectives that are vague and constraints that narrowly only focus on the total cost of ownership (TCO) [2]. In the case of a CMMIS, an example of this scenario can be setting the objective as "having data on equipment condition and failure alerts" and a TCO constraint set at some maximum cut-off value. In best cases, the objectives require specific metrics such as increase in production and reduction in costs, equipment downtime, and labor utilization, accompanied by detailed cost constraints on the various aspects of the system design and implementation project, complemented by a clear definition of organizational project strategy such as "to achieve lean operations", "minimum operations costs", etc..

However, even such a best case is not holistic. If the system is required to address other non-financial and critical issues in addition to the financial criterion, the objectives, constraints, and requirements strategy should be based on criteria such as safety, environment, risks, etc. as well. Note that objectives, constraints, and strategy can be based on one or more of these criteria. For example an organization may choose to set financial, environmental, and safety objectives; With economic, environmental, risk, and time constraints; And a strategy based on projects economics, availability of options, and time sensitivity.

2.1.3 Maintenance Management Method Assessment

This component is to distinguish the special requirements of a CMMIS related to the required maintenance management method and practice for a specific application.

This includes the required system logic, which in this context refers to the rules, processes, methods, procedures, models, and practices of the maintenance management system and the organization's maintenance strategy. It also includes any ontology or standards which the practice is based on, or more specific functional standardized specifications such as ISO 13374-1 for machinery diagnostic systems.

This component also includes any required general or application specific decision analysis method, and assessment of design, development, and deployment methods.

2.1.4 Resources Assessment

In general, the resource categories are financial capital, human capital, information, and technology. Other tangible and non-tangible assets may be included under information and technology categories.

2.1.5 Selection and Definition of Performance Criteria

In order to design an effective system, and to be able to perform meaningful performance evaluation after its implementation, the framework calls for designation of the criteria that are critical to the organization and which are expected to be addressed by the CMMIS when performing the needs assessment activities.

Once the performance criteria are determined, suitable metrics can be proposed and defined for each criterion. The criteria and the defined metrics would then guide system design, implementation, and performance measurement activities.

2.2 Performance Criteria and Metrics Development for Performance Evaluation

Once the CMMIS system performance criteria are established by the needs assessment activities, the performance criteria development activity of the framework is intended to explore these criteria as they relate to the organization's objectives and business operations, and to select or develop metrics for each criterion so that effective design and performance measurement could be achieved. During this phase, the priority of the criteria, their relationships, and how they relate to the organization's strategy would be developed. Note, that the framework calls for development of performance criteria and their metrics prior to design, so that the design phase can be goal oriented and the expectations for performance can be clarified and laid out in advance. Several performance criteria and their respective metrics for system performance measurement are discussed in this section.

2.2.1 Maintenance, Reliability, and Availability

The key objective of a CMMIS is to improve a maintenance and reliability engineering practice. Performance measurement of a CMMIS with respect to these criteria has been studied extensively ([3] – [12]) In addition to some industry specific maintenance performance indicators mentioned by [12], the following performance metrics related to achieving effective maintenance are typical:

- increase in the detection of processes not performing within specifications;
- higher serviced items per maintenance downtime; and
- decrease in production delays due to untimely outages or maintenance issues.

2.2.2 Financial

Financial metrics have traditionally been one key performance criterion for assessing the value of CMMIS functionalities. Although it is sometimes difficult (or meaningless) to translate all aspects of performance in terms of monetary figures, if done properly, financial figures can be used to more readily calculate and demonstrate the value of investments and the value of various

system performance levels. Some areas that traditionally financial metrics have been used are:

- improvements on operations cost savings;
- increase in sales due to production increase; and
- decrease in maintenance costs and cost overruns.

These metrics would typically be expressed by discount cash flow (DCF) methods.

2.2.3 Environmental

Organizations face larger internal and external pressures to reduce the environmental impact of their operations. This concern extends to the CMMIS, where there is great value in assessing the performance of the system based on environmental metrics ([3] and [13]).

2.2.4 Safety

Safety is perhaps the most important consideration in business decisions, and hence should be one of the main performance criteria of a system. Safety has been one of the key performance criteria in the work of many maintenance performance measurement practices ([3] and [15] – [17]). Metrics that can measure how the system improves safety are greatly valuable to an organization, and can include:

- potential for decrease in false negative alerts; and
- potential for increase in overall safety alertness level.

2.2.5 Information Flow

Business value creation in the modern economy is information driven, however, possession and storage of information does not mean that it would be effectively used to create value. The most fundamental function of a CMMIS is to provide effective information flow, from the point where data and information is generated, to where it is retrieved and used, including all the provisions for storing, preserving, and securing it. Metrics for measuring the effectiveness and efficiency of information flow in CMMIS can be related to:

- ability to collect and store more asset, equipment, and system data as needed, i.e. be appropriately scalable; and
- increase in the quality and better conditioning of data.

2.2.6 Risk Awareness

All businesses face various types and levels of risk. An effective CMMIS is able to cast light on the sources and types of risks that can threaten an organization. This makes CMMIS particularly valuable since it can monitor types and levels of risks arising from critical assets and processes which an organization is most sensitive to.

Risks and uncertainties facing an organization are related, and many of the metrics related to the risk awareness performance of a system have much in common with metrics related to system performance in uncertainty handling (next section). In this text the difference between uncertainty and risk is that although both can be expressed with the probability axioms, risk has a negative connotation, and is a subjective term. For assessing system performance in providing risk awareness, some specific metrics can be related to the accuracy and correctness of:

- impact of risk on costs, schedule, and production;
- risk uncertainty estimation and quantification of subjective risk estimations ([18], [19]); and
- risk classification and correlations with other risks;

2.2.7 Uncertainty Handling

An effective CMMIS is driven by effective information flow, and thus affected by information uncertainties, i.e. the system not only relies on effectively handling information and its flow in the organization, but also on uncertain knowledge and judgement of experts and decision makers, and sources of information uncertainty.

Many provisions and functionalities that can increase the effectiveness of a CMMIS in handling uncertainty are common with the ones described as part of the risk performance criterion, which were mainly related to improving uncertainty identification, quantification and estimation. However, not all uncertainties are related to risk events. The metrics related to a CMMIS in handling uncertainty can be related to measuring improvements on:

- identification of the areas and factors of uncertainty;
- understanding how these uncertainties are quantified and expressed;
- clarifying the relation between uncertainties and business risks; and
- understanding the relations and correlations between uncertainties and propagation of uncertainty from equipment and maintenance floor to high level performance indicators;

2.2.8 System Embedded Intelligence and Automation

A distinguishing feature of an effective CMMIS is its value creation capability by providing system intelligence, analytics, and automation. With significant advances in machine learning and data mining algorithms and software, and facing large amounts of data that require the decision maker's attention, an intelligent CMMIS can provide greater value by, e.g., enabling knowledge discovery from large amounts of data, employing models learned from data for reliability and failure predictions, and allowing for online anomaly detection ([20] – [26]).

Performance of prediction and control models, algorithms, or methods are typically measured by observing the deviation of their outputs from a true outcome, i.e. their error. So a typical metric for performance is prediction and control accuracy. On the other hand, other tangible metrics that translate to business value can be:

- reduction in human activity time on tasks;
- increase in human value-added activity; and
- decrease in instances of human intervention in tasks.

2.2.9 Operational Excellence, Strategic Directives, and Quality

Operational excellence may be an integral part of an organization's management system and business model, and hence suitable metrics are required to gauge the performance of the CMMIS in this regard. Such metrics could be related to lean operations and continuous quality improvement principles. Metrics related to production operations could involve measuring the role of CMMIS in production increase and expansion. Related to strategic advantages, metrics could involve product time-to-market and customer satisfaction. Quality of process and operational excellence in maintenance has been an integral part of the total quality management practice in operations, and addressed by many practitioners and scholars. e.g. in [10], and [27] – [30].

2.2.10 Knowledge Management

Another aspect of CMMIS performance measurement is its capability in providing knowledge management [31] and the

effectiveness of its knowledge base system. Metrics in this regard could be related to:

- assignment of process and business context to data;
- inclusion of meta data associated with data;
- knowledge extraction from unstructured data;
- capturing tacit knowledge in processes; and
- effective visualization of knowledge.

2.2.11 Decision Quality

An effective CMMIS enables good decision making at various levels of the organization. All the discussed aspects of a CMMIS so far are required to work in concert to enhance the decision making process, and to improve decisions.

The extent to which a CMMIS improves the decision making process and increases the quality of decisions is itself an important performance criterion for consideration. Many factors affect decisions and decision making, from the ability to technically analyze decisions as problems with multiple attributes, trade-offs, and uncertainties ([32] and [33]), to cognitive issues in human decision making ([34] and [35]).

For example, performance metrics related to a CMMIS affecting decision quality could be considered in relation to how the system handles uncertainty and provides risk awareness, e.g. whether right questions have been asked to assign the right subjective uncertainties to inputs, or if there have been proper understanding of relative probabilities and conditional probabilities. Furthermore, consistency in use of quantification and decision making methods can be measured as a metrics for decision quality.

2.2.12 Decision Options and Flexibility

There usually exists flexibility and options in making decisions, such as delaying, modifying, upgrading, reversing, or abandoning actions, where such flexibility has itself intrinsic value. The performance of the CMMIS in allowing for flexibility and having options in decision making could be measured by comparing it with the case where such flexibility does not exist.

The advantage of having options in adding features to the CMMIS and also decision making for the users of the CMMIS is valuable and may not necessarily exist [36], hence it is also important to measure the performance improvement due to flexibility and options in a maintenance process.

2.2.13 Organizational Adaption and User Interaction

Many technology products and systems fail to get adopted and used by the intended users, despite careful considerations and sophisticated efforts in designing, engineering and developing the technology.

This issue is rarely related to the technical sophistication of the system, and has been shown to be mainly sociological, organizational and psychological, discussed in depth in [37] and [38].

2.2.14 System Security

Information security is among the most important considerations in designing and implementing a CMMIS. System security performance has been researched extensively. Some commonly used metrics related to performance measurement of IIS security in general are:

- decrease in instances of intrusion;
- increase in the frequency of intrusion detections;

- decrease in activities related to arranging and dealing with authentication issues; and
- decrease in the amount of data loss.

2.2.15 Presentation of Information

Effective performance of a CMMIS not only depends on how the system collects and processes information and facilitates its flow, but also on how it presents relevant information to various users of the system. Many information systems have failed to provide their intended level of performance due to inadequate or ineffective presentation of information and user interaction despite sophistication in their functionality ([37] and [39]). Performance metrics for presentation and reporting in the context of a CMMIS, can be related to:

- decrease in the amount of time spent on training users;
- increase in user ratings related to using the system and understanding information presented;
- timely alerting related to sensitive productivity or quality loss;
- clear indication and alerting of safety risks and environmental hazards; and
- reporting and alerting in the case of non-compliance with regulations.

2.2.16 Organizational Learning

Improvement in a system occurs when a learning mechanism is present. Learning occurs by comparing outcomes to ideals, or by comparing two consecutive outcomes.

Achieving continuous organizational learning and improvement of the organization's performance based on past data, decisions, and actions within a CMMIS is perhaps the most difficult level of performance to reach. Effective organizational learning builds on effective system performance based on many of the criteria that was previously discussed.

2.3 Performance Criteria Priority

Once the criteria for CMMIS performance evaluation are established by the organization, the priority of the criteria, their relationships, and how they relate to the organization's strategy should be developed. The organization may broadly categorize its required performance criteria in terms of implementation priority or urgency, e.g. by a priority pyramid as illustrated in Figure 3. In this model, the right arrows indicate the ascendancy in the pyramid building blocks, i.e. decreasing priority but increase in importance. The left arrows indicate that the top portions affect the lower parts of the pyramid, implying that although the top categories may be rated with less urgency (or priority) than the foundation criteria, but they are the most important in terms of achieving an effective system with high performance.

Note, that the particular ranking of criteria in Figure. 3 is illustrated only as an example, and ultimately, each organization should set and assess its own performance criteria and rank their priority and urgency based on factors that are intimate to the organization. This requires careful consideration and balancing of organizational needs at different times and situations. Moreover, when the relation between choices and the organization's strategy is under consideration, the balanced score card approach can be used [40]. Choices can be evaluated and ranked with more quantitative methods such as the analytical hierarchy process (AHP) or PROMETHEE.

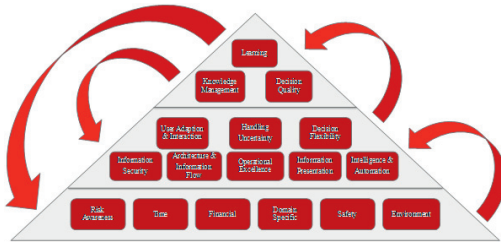


Figure 3. Criteria ranking: the performance criteria pyramid

3. A VERIFICATION PROCESS FOR ASSESSING THE EFFECTIVENESS OF THE FRAMEWORK

One approach for verifying the effectiveness of the framework is using case studies. With this approach, effectiveness of the framework as a hypothesis will be confirmed when potential for improved system implementation after applying it to cases can be demonstrated [41].

The case study could entail applying the framework to design a CMMIS for a certain organization to demonstrate how the framework can clarify the value proposition of a CMMIS based on the organization's needs and required performance criteria, and how it could result in a fit-for-purpose and performance-based system that is superior to the existing practice of the case, based on relevant metrics.

As a result of applying the framework, a hypothetical CMMIS would be designed for the case. Simulated performance of the hypothetical CMMIS could then be compared with the existing practice of the case organization. These comparisons could be conducted according to select performance criteria metrics, and by performing simulations (where required) on workflow, information flow, decision making based on decision options, and system business dynamics, based on the case operations. The comparisons could be performed to measure the degree to which the framework performs better than the existing practice, and its effectiveness in resulting in a fit-for-purpose, holistic, and performance-based system.

Simulations on work flow and information flow would simulate case operations, production, occurrence of fault and failure events, fault detection and prediction, maintenance activities, human and system decisions, task scheduling, and other discrete events. A range of distributions for the input variables could be considered representing variable uncertainty, resulting in a Monte Carlo type simulation. Simulation output would then be expressed based on relevant performance metrics.

4. CONCLUSIONS

In this work, a framework was introduced for effective implementation of CMMIS systems. The framework was developed with the objective of achieving a fit-for-purpose, holistic, and performance based system. Concretely, it was noted that a performance-based system is when the benefits due to the system are balanced with associated costs. A fit-for-purpose system is designed with respect to organizational business strategy, objectives, and constraints. And a system is holistic when it addresses various performance criteria.

The importance of IIS for a CMMIS was mentioned. It was emphasized that the framework takes an IIS centric view for the design of a CMMIS, and has been developed considering an organizations enterprise system.

The two key phases of the framework, i.e. the needs assessment and the performance criteria and metrics development phases were discussed. This included the main aspects of the needs assessment phase, i.e. existing system assessment, requirements assessment, maintenance management method assessment, and resources assessment.

The performance criteria and metrics development phases were also discussed. Some performance criteria for holistic performance assessment of maintenance systems were specifically mentioned:

1. Maintenance, reliability, and availability performance;
2. Financial
3. Environmental
4. Safety
5. Information flow;
6. Providing risk awareness;
7. Handling uncertainty;
8. System embedded intelligence and automation;
9. Operational excellence, strategic advantages, and quality;
10. Knowledge management;
11. Decision quality;
12. Decision options and flexibility;
13. Organizational adaption and user interaction;
14. System security;
15. Presentation of information; and
16. Organizational learning.

Each criterion was described as it relates to the performance of a CMMIS. Furthermore, for each criterion, sample metrics were mentioned that could be used for specific performance measurement.

While many performance criteria exist, the need for prioritizing them and their relative metrics, including a simple and visual criteria ranking method, were briefly discussed.

In the end, a procedure for verifying the effectiveness of the framework in achieving its objectives was laid out.

5. ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support of the national science and engineering research council of Canada (NSERC).

6. REFERENCES

- [1] R. L. Van Horn, A. B. Schwarzkopf and R. L. Price, *Information systems solutions, a project approach*, New York, USA: McGraw-Hill, 2006.
- [2] B. G. Ferrin and R. E. Plank, "Total cost of ownership models: an exploratory study," *Journal of Supply Chain Management*. vol. 38, no. 3, pp.18-29, 2006.

- [3] P. N. Muchiri, L. Pintelon, H. Martin, and A.-M. De Meyer, "Empirical analysis of maintenance performance measurement in Belgian industries," *International Journal of Production Research*, vol. 48, no. 20, pp. 5905–5924, Oct. 2010.
- [4] A.K.S., Jardine and A.H.C., Tsang, *Maintenance, Replacement, and Reliability*, Boca Raton, FL: Taylor and Francis, 2006.
- [5] J., Moubray, *Reliability Centred Maintenance*, 2nd ed., Oxford, UK: Butterworth-Heinemann, 1997.
- [6] J.D., Campbell, *Uptime, Strategies for excellence in maintenance management*, Portland, OR: Productivity Press, 1995.
- [7] A.H.C., Tsang, A.K.S., Jardine, and H., Kolodny, "Measuring maintenance performance: a holistic approach," *International Journal of Operations & Production Management*, vol.19, no.7, pp.691–715, 1999.
- [8] W.J., Moore and A.G., Starr, "An intelligent maintenance system for continuous cost-based prioritisation of maintenance activities," *Computers in Industry*, vol.57, no.6, pp.595–606, 2006.
- [9] C.A., Bana e Costa, M.C., Carnero and M.D., Oliveira, "A multi-criteria model for auditing a Predictive Maintenance Programme," *European Journal of Operational Research*, vol.217, no.2, pp.381–393, 2012.
- [10] M., Ben-Daya and S.O., Duffuaa, "Maintenance and quality: the missing link", *Journal of Quality in Maintenance Engineering*, Vol.1 No. 1, pp.20-6, 1995.
- [11] J.M. Simoes, C.F., Gomes, and M.M., Yasin, "A literature review of maintenance performance measurement," *Journal of Quality in Maintenance Engineering*, vol. 17, no.2, pp.116-137, 2011.
- [12] A., Parida and U., Kumar, *Handbook of maintenance management and engineering*, Ben-Daya, M., Duffuaa, S.O., Raouf, A., Knezevic, J. & Ait-Kadi, D., Eds., London, UK: Springer-Verlag, 2009.
- [13] R.M.C. Ratnayake, and T. Markeset, "Technical integrity management: measuring HSE awareness using AHP in selecting a maintenance strategy," *Journal of Quality in Maintenance Engineering*, vol. 16, no.1, pp. 44-63, 2010.
- [14] K. Y. Kutucuoglu, J. Hamali, Z. Irani, and J. M. Sharp, "A framework for managing maintenance using performance measurement systems," *International Journal of Operations & Production Management*, vol. 21, no. 1/2, pp. 173–195, Jan. 2001.
- [15] C. K. H. Hon, A. P. C. Chan, and D. W. M. Chan, "Strategies for improving safety performance of repair, maintenance, minor alteration and addition (RMAA) works," *Facilities*, vol. 29, no. 13/14, pp. 591–610, 2011.
- [16] V. Narayan, "Business performance and maintenance: How are safety, quality, reliability, productivity and maintenance related?," *Journal of Quality in Maintenance Engineering*, vol. 18, no. 2, pp. 183–195, May 2012.
- [17] C. K. H. Hon, A. P. C. Chan, and D. W. M. Chan, "Strategies for improving safety performance of repair, maintenance, minor alteration and addition (RMAA) works," *Facilities*, vol. 29, no. 13/14, pp. 591–610, 2011.
- [18] I.N., Durbach, and T.J., "Stewart, An experimental study of the effect of uncertainty representation on decision making," *European Journal of Operational Research*. no.214, pp.380-392, 2011.
- [19] E., Payzan-LeNestour and P., Bossaerts P., "Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings", *PLoS Computational Biology*. vol.7, no.1: e1001048, 2011.
- [20] S., Dash and V., Venkatasubramanian, "Challenges in the industrial applications of fault diagnostic systems," *Computers & Chemical Engineering*, vol.24, no.(2–7), pp.785–791, 2000.
- [21] B.C., Juricek, D.E., Seborg, and W.E., Larimore, "Predictive monitoring for abnormal situation management," *Journal of Process Control*, vol.11, no.2, pp.111–128, 2001.
- [22] V., Venkatasubramanian, R., Rengaswamy, K., Yin, and S.N., Kavuri, "A review of process fault detection and diagnosis: Quantitative model-based methods," *Computers & Chemical Engineering*, vol.27, no.3, pp.293–311, 2003.
- [23] C., Emmanouilidis, E., Jantunen, and J., MacIntyre, "Flexible software for condition monitoring, incorporating novelty detection and diagnostics," *Computers in Industry*, vol.57, no.6, pp.516–527, 2003.
- [24] V., Uraikul, C.W., Chan, and P., Tontiwachwuthikul, "Artificial intelligence for monitoring and supervisory control of process systems," *Engineering Applications of Artificial Intelligence*, vol.20, no.2, pp.115–131, 2007.
- [25] A., Chan, and K.R., McNaught, "Using Bayesian networks to improve fault diagnosis during manufacturing tests of mobile telephone infrastructure," *Journal of the Operational Research Society*, vol.59, no.4, pp.423–430, 2007.
- [26] M., Hajizadeh, and M.G., Lipssett, "Anomaly detection in mining haul truck suspensions using wavelet analysis of strut pressures and parametric modeling of strut damage," *Proc 8th Int'l Conf on Condition Monitoring and Machinery Failure Prevention Technologies, CM 2011 and MFPT 2011*, 2011.
- [27] A. H. C. Tsang, "Strategic dimensions of maintenance management," *Journal of Quality in Maintenance Engineering*, vol. 8, no. 1, pp. 7–39, Jan. 2002.
- [28] S. Nakajima, "Introduction to TPM," New York, NY: Productivity Press, 1988.
- [29] J.M., Sharp, Z., Irani, T., Wyant, and N., Firth, "TQM in maintenance to improve manufacturing performance," *Proceedings of PICMET conference*, 1997.
- [30] P., Willmott, "Total Productive Maintenance: The Western Way," New York, NY: Butterworth-Heinemann, 1994.
- [31] T.H., Davenport, and L., Prusak, "Working knowledge, how organizations manage what they know," Harvard, MA: Harvard Business School Press, 2000.
- [32] R.A., Howard, "Decision analysis: practice and promise. *Management Science*, vol.34 no.6, pp.679–695, 1988.
- [33] R.L., Keeney, and H., Raiffa, *Decisions with multiple objective, preferences and value tradeoffs*. Cambridge, UK: Cambridge University Press, 1993.

- [34] A., Tversky, and D., Kahneman, "Judgement under uncertainty: heuristics and biases," *Science*. vol.185, no.4157, pp.1124-1131, 1974.
- [35] B., Schwartz, "Tyranny of choice. *Scientific American*," vol.290, no.4, pp.70-76, 2004.
- [36] L.C., Wu, L.H., Wu, and Y.F., Wen, Y.-F., "Interdisciplinary research of options theory and management information systems: Review, research issues, and suggestions for future research," *Industrial Management & Data Systems*, no.110, pp.433-452, 2010.
- [37] D.R., Dixon, D.R., "The behavioural side of information technology," *International Journal of Medical Informatics*. vol.56, no.(1-3), pp. 117-123, 1999.
- [38] E.M., Rogers, "*Diffusion of Innovation*," 5th Edition. New York: Simon & Schuster, 2003.
- [39] S., So, and M., Smith, M., "The impact of presentation format and individual differences on the communication of information for management decision making," *Managerial Auditing Journal*, vol.18, no.1, pp.59-57, 2003.
- [40] R.S., Kaplan, and D. P. Norton, "Using the balanced scorecard as a strategic management system," *Harvard Business Review*, January-February, 1996.
- [41] A. Carter-Steel and Al-Hakim, *Information systems, research methods, epistemology, and applications*. Hershey, PA, USA: Information Science Reference, 2009.

Usability aspects of eMaintenance solutions

Karina Wandt

Division of Operation, Maintenance
and Acoustics

Luleå University of Technology
Luleå, SE-971 87, Sweden
+46 920 49 24 83

karina.wandt@ltu.se

Phillip Tretten

Division of Operation, Maintenance
and Acoustics

Luleå University of Technology
Luleå, SE-971 87, Sweden
+46 920 49 28 55

phillip.tretten@ltu.se

Ramin Karim

Division of Operation, Maintenance
and Acoustics

Luleå University of Technology
Luleå, SE-971 87, Sweden
+46 920 49 23 44

ramin.karim@ltu.se

ABSTRACT

As technological development progresses in society, there are several new possibilities to make the maintenance process in industry more efficient. Today, eMaintenance solutions facilitate data management in maintenance activities; data can easily be integrated and shared between sources in heterogeneous environments. This enables systems used in maintenance, such as Computerized Maintenance Management Systems (CMMS), to base decisions on various data, e.g. data produced in other processes. These systems often fulfil the technical requirements (e.g. data consistency control) required to support activities in the maintenance process (e.g. management, support planning, and assessment), but the human interaction with the systems is still essential to the quality of the work performed. Since many maintenance activities require manual input of data the interaction between user, e.g. technician, and system, e.g. CMMS, has impact on the quality of the data; in order for the data to be right and relevant, the technician may need supporting directives from the CMMS. Hence, the system usability must be considered when assuring the quality of the manually inputted data.

The focus of this paper is to investigate CMMS limitation issues due to usability aspects. Furthermore, the paper discusses the role of context awareness within user interfaces managing data obtained through eMaintenance solutions and presents ideas for future research on context awareness in eMaintenance solutions. Data and conclusions have been conducted through literature studies and case studies within the area.

Keywords

CMMS, Context awareness, motivation, data quality.

1. INTRODUCTION

Technology development is in constant progress and the tools and systems used in maintenance processes are continuously improving. As the tools and technologies get more advanced, the need for high information processing solutions with that can communicate between systems is increasing [14]. Today, ICT-based maintenance solutions, i.e. eMaintenance solutions, support enterprises with effective and efficient decision-making by enabling just-in-time access to maintenance information [13][26]. eMaintenance solutions aim to support maintenance stakeholders with necessary information adapted to their context [14]. A proper eMaintenance solution needs to be able to sense the context of the individual stakeholder in order to properly adapt the information to the stakeholder's current situation [3]. eMaintenance solutions with context-sensing capabilities facilitate the interaction between the system, e.g. Computerized Maintenance Management System

(CMMS), and its specified users, which in turn will contribute to effectiveness and efficiency in the maintenance process. The increased availability of data enables accurate and precise decision-making in maintenance, given that the collected data are relevant, used correctly and hold the expected level of quality. The quality of the data is essential in the decision-making process, since inaccurate or poor data may influence the decision incorrectly, especially if the system using the data is not able to verify the quality [24]. Hence, it is important for maintenance tools that are using eMaintenance solutions to be aware of data quality issues and actively reduce them [16][24].

There are many factors that can affect the quality of eMaintenance data and various studies [1][15] have emphasized the importance of the users' interaction with the system. A poorly designed system can actually promote human errors when using the system [19][21]. A system in this paper is defined as the computer programs used for surveillance, storage and analysis of operation and maintenance data. As systems become more advanced, due to e.g. increased automation and greater complexity, it gets harder for the users (or system operators) to keep track of the actions in the system and to fully understand the system capability [21]. A CMMS is most likely capable in executing the task it has been designed for, but it often lacks the capability to assist the user to perform the task, resulting in data quality issues and other criticism [1].

The purpose of this paper is to further investigate the usability of a CMMS and the paper seeks to address the following question: What are the main usability issues within a CMMS and what are the core reasons for these issues?

2. THE INTERACTION BETWEEN USER AND EMaintenance SOLUTIONS

A typical interface between eMaintenance solutions and the user of eMaintenance data is a CMMS. A CMMS organizes many different functions, such as work orders, inventory control, asset management and integration towards other management systems. Operators, technicians, as well as, management may be typical users of a CMMS [15]. A CMMS provides work orders to plan and schedule inspections, preventive maintenances and maintenance of machine break-downs; this includes assigning personnel, reserving materials, recording costs and tracking information history. When appropriately configured, and interfaced, with other systems within a company, such as an Enterprise Resource Planning system (ERP), a CMMS can become a critical and useful tool in handling maintenance organization activities [17]. The capabilities of handling large

quantities of data facilitate a thoughtful maintenance approach [15]. However, the ability of handling large amount of data is not much worth if the data is not used correctly. The term black holes in CMMS was introduced by Labib [15] as a description of systems greedy for data input that seldom provide any output in terms of decision support. In general there is a lack of basic user-friendly design in many CMMS and this is a result of the systems being designed for accounting and/or IT rather than for the users specific needs [15].

2.1 System usability

The awareness of the importance of system usability is constantly increasing and the European Parliament and the council of the European communities has established a directive that requires employers to ensure usability when designing, selecting, commissioning or modifying software. The following requirements are listed in the directive; (i) the software must be suitable for the task; (ii) the software must be easy to use and, where appropriate, adaptable to the user's knowledge or experience; (iii) the systems must provide feedback on performance; (iv) the system must display information in a format and at a pace that is adapted to the user; and (v) the system must adapt to the principles of software ergonomics [5][6].

Since a CMMS is very complex, containing information of the whole production process, such as, financial information, logistics, parts and services, maintenance schedules, etc., it is not a simple system to build. Because of its complexity there is a risk in the system becoming error prone, especially for the users. Therefore it is important that the system users are active in the design of the system. Rather than being the cause of an error; "operators tend to be the inheritors of system defects created by poor design, incorrect installation, faulty maintenance and bad management decision" [21]. Therefore it is important that the designers of complex systems need to assume that every mishap will happen and their job is to design against it. To assist in this three principals are recommended [19] for the design of complex systems:

1. Use common knowledge, found in the real world, for actions to be used in the operation of the system. Usage should not require unique knowledge to operate.
2. Use functions that help the user to make natural decisions based on both natural and artificial constraints.
3. Make options readily visible to execute actions and make the results of each action readily available for evaluation.

To make a system easier to learn and use is to allow the users to explore, experiment and learn different possibilities of the system. To make a system explorable and decrease the risk for user errors it needs to: (i) allow for the user to know what state they are in the system and readily see what they need to do, (ii) the effect of the action must be both clear and easy to understand and (iii) undesirable actions should be reversible [19]. This is necessary to increase the usability of complex systems, e.g. CMMS.

A system with a high level of usability should be easy to learn, engaging to use, and support the user to efficiently and effectively complete tasks and goals [25]. ISO 92441 describes usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [12]. This definition includes objective measures, such as specified user context and

the goal, of how usable an application is. Goal-oriented actions are often connected to motivation and in order to help the user to achieve a specific goal, the goal must be clear and the system should be able to motivate the user. Having the usability definition in mind, the following two sections discuss the impact of user motivation and context awareness.

2.1.1 The impact of user motivation

The usability of a CMMS is a challenge and it is important to consider the users' level of motivation to know how to work in improving system weaknesses like usability. Motivation can be divided into autonomous extrinsic motivation, which is based on importance and outer goals, and intrinsic motivation, which is based on interest and inner goals. Both types are related to the performance, satisfaction, trust and well-being at the workplace [8][27]. Autonomy, stimulating tasks and decision authority are important aspects to consider when talking about work motivation. The job characteristics model, illustrated in Table 1, is based on the idea that the task itself is the key to employee motivation. The model identifies five characteristics and the fundamental principle of the model is that engagement derives from inside in order for the work to feel satisfying and self-rewarding.

Table 1. The job characteristics model

Characteristic	Degree of...	Description
Skill variety	...skills required.	Describes different skills and talents needed for a given task.
Task identity	...clarity of task definition.	Clear definition of beginning, middle, end and visible results.
Task significance	... "Substantial impact".	Describes how meaningful a given task is.
Autonomy	...freedom.	Includes the ability to schedule work, as well as, finding a solution for a given task.
Job feedback	...relevant feedback.	Keep the employee informed about the performance.

Pride of being responsible and feeling ownership with one's work is also an important aspect to be considered when dealing with motivation. It is also true that a sense of ownership of the equipment affects how one treats the equipment and that maintenance will directly affect its reliability and performance. As employees develop this sense of ownership there is an increased sense of pride, motivation and self-esteem. The long-term impact is increased productivity [7][18]. When the users feel alienated to the equipment they will not be motivated to exert themselves in improving the equipment nor taking any extra steps to maintain them.

2.1.2 The impact of context awareness

A necessary, yet difficult, task for system design is the aspect of context awareness. As mentioned before, to make a system easier to use one should know where they are in the system. The system should also be able to allow each user to access relevant information and actions. The CMMS needs to provide users involved in a certain context of the maintenance process with functionality providing accurate maintenance-related information at the right time and adapted to the actual maintenance context

information at the right time adapted to the actual maintenance context [14]. The context is any information that can be used to characterize the situation of an entity [4][26], wherewith the entity can be a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves [19]. Context awareness should then use the context template to provide task-relevant information and/or services to a specific user [4][26].

Thus context awareness can be described as the ability of a system to adapt the operations to the current context without explicit user intervention [2][23] and thereby respond to changes in the environment in order to make the system behave more relevant to the current situation. The aim for systems with context awareness is to increase usability and effectiveness by taking the actual situation and the respective user context into account [2].

The actual context or environmental context can be divided into an external and internal dimension [9][11][19]. The external dimension handles attributes, such as location, light, sound, movement, touch or temperature, which can be measured by hardware sensors. The internal dimension handles attributes that is mostly specified by the user or captured by monitoring user interactions, e.g. the user's goals, tasks, work context, business processes or the user's emotional state.

3. METHOD

Two case studies were conducted in order to find the core reasons for usability issues within a CMMS. The base for the study was the usability definition provided by ISO 92441 [12] described in section 2.

The first case study was conducted to gain insight into the main usability issues of the CMMS, while the second case study was conducted to find out how the CMMS could be improved. Semi-structured interviews were chosen for both case studies since the goal of the studies was to "pick up" as many different ideas concerning the CMMS as possible without hindering the interviewees from speaking freely about the system.

The interviews of the first study were conducted at several companies; two mining, five pulp and paper and an airplane manufacturer. The goal of this case study was to gain a more complete understanding of how users of the CMMS find the system helpful in their work. The 10 individuals interviewed had the following roles in the organizations: section manager, foreman for mechanical maintenance, maintenance mechanic, foreman for electrical maintenance, electrical maintenance mechanic, maintenance engineer, maintenance technician, maintenance mechanic for preventative maintenance, production operator and an automation systems designer. The interviewed were experienced and had a good understanding of the system.

The second case study was conducted to gain more specific information concerning the usability of the CMMS and how it could be improved. It was conducted as a group interview with a team of four aircraft service personnel who had different roles in using the CMMS. This was done to learn as much as possible from the individuals on how the whole system functions. The questions used in the open-ended interview were: "How easy is it to use the CMMS?", "How does one get help if needed?", "What limitations are the most frustrating?" and "What improvements are recommended?".

The interview responses were recorded and analysed. For the first case study, the results were categorised in accordance to the types of problems the respondents brought forth. In the second case study, the responses were categorised based upon the research questions used. Everything was controlled and compared to the original material.

4. RESULTS

The two case studies found in this paper look at user issues of the CMMS in several areas and this section presents the collected data.

4.1 Case study 1

The following problems were identified as issues relevant to the ICT solutions in maintenance activities.

(1) *Lack of compatibility*: The CMMS was not connected with production, material management or spare parts storage and purchasing. These together are required for operative and strategic decision making in the maintenance phases. One reason for this problem is that the data were not comparable and difficult to combine. This complicated maintenance related operative and strategic decision-making.

(2) *Data Multiplicity*: The CMMS did not collectively store technical documentation and drawings. Certain information regarding technical documentation and drawings was sometimes stored in multiple sources. Since some of the stored information was not regularly updated or maintained it was difficult to know if it could be trusted and sometimes the sources also contained conflicting information.

(3) *Manual data input and transfer*: The CMMS required manual input and transfer of data into the system. Manual data entry resulted in data of varying quality; some personnel enter detailed and descriptive failure reports, while others unknowingly omit compulsory data. Some automation based upon work orders, production stops, etc. could also have been helpful.

(4) *Not user friendly*: The CMMS user interface and the systems functionality were not intuitive. Several of the participants shared the opinion that the usability of the maintenance systems could be improved. At times it could be difficult to find relevant information and the user guidance was lacking.

(5) *Lack of guidance*: The CMMS lacked user instructions that were easy to use. There was no common practice for what kind of information that was needed and by whom and in what form it should be collected in the different phases of the maintenance process. Several different methods to collect data existed and, therefore, information in the database was not compatible. This made it difficult to do systematic data analysis and to produce key performance indicators.

(6) *Limited strategic use of CMMS*: The CMMS data were not generally used for operative and strategic decisions. Thus, the CMMS was mainly used for information transfer and communication between people of different positions and organizations, e.g. sending and receiving work orders or accessing information from the system. Since the data were not collected for operative and strategic decision making, the CMMS users were not always aware of what purpose the data were to be used for and the user motivation to collect and enter good quality data in the system was thereby lowered.

(7) *Low competence*: The CMMS was not used proficiently by the people it was intended for because only a few were able to use the system. Even though the maintenance personnel mainly used the CMMS, the operative personnel conducted a lot of maintenance actions but did not have access to the CMMS. This led to the situation that only part of the actual maintenance actions were recorded in the CMMS and a great deal of data was lost. Finally, when the operative and strategic decisions were made, they were based on the recorded data, which only showed a part of what had occurred. Therefore, the strategic decisions were often not well grounded.

4.2 Case study 2

The following CMMS issues were identified in this case study.

(1) *Better training needed*: No course was held on the CMMS, all training needed was performed online as a self-study course prior to the introduction of the CMMS. This placed a lot of responsibility on the user to plan and prioritize the training. Due to this the user acceptance of the CMMS was thereby lowered.

(2) *Informative feedback is needed*: The CMMS did not produce informative feedback to the user. The CMMS should clarify the action and should not only say e.g. "Wrong input data, keyword is missing". A suggestion was to add feedback that will explain the purpose of the keyword.

(3) *Orientation within the system is needed*: The feeling of being a part of a whole process was somehow lacking and the significance of the work performed could therefore sometimes be questioned. The following information was pointed out as important for the case study participants:

- "Visualization of actions before and after me in the chain."
- "Visualization of the importance of my effort and my part in the process."
- "Visualization of what has been done, not only what is about to come."
- "Visualization of the number of completed actions of the team. (Important that this is not presented as a stress factor, but as positive feedback.)"

(4) *Intuitive Interface is needed*: Many minor interface restrictions created irritation. The CMMS interface was not intuitive and did not help the user to find the most likely action in a specific situation. The following examples were mentioned:

- "Need to click twice in a text field before editing."
- "Little or no feedback from the system."
- "It is not possible to switch text field with the arrows, must use TAB."

(5) *Flexibility*: The work process was not always aligned with how the CMMS was designed, e.g. there was no natural or consequential process. A lack of flexibility in the system also forced the users to develop work arounds.

(6) *Simplify common actions*: Necessary work arounds ended up with time consuming phone calls with other departments to solve issues due to that the CMMS did not support actions that are commonly used by the operators.

(7) *Interface should reflect end-customer solutions*: The CMMS was implemented in 2010+, but the interface was a reminder from the mid 90's. Many of the actions required and solutions given to the users did not reflect common interfaces. This resulted in lowering the users' motivation to use the system.

5. DISCUSSION

The results of the two studies were compiled and analysed based upon the seven groupings from case study 1. Beginning with Lack of compatibility, data multiplicity, manual data input and transfer, not user friendly, lack of guidance, limited strategic use of CMMS and, finally, low competence.

First, the lack of compatibility of the interface showed a need for familiarity in the interface and improved CMMS functionality. The interface should reflect common end-user solutions of today, both mobile and stationary computer operative systems have standardized order of operations, in which, users can identify with. Therefore the proposed solution would be to improve the user interface so that it reflects commonly used solutions.

Secondly, data multiplicity pertains to the patchwork of solutions used to help and guide the user when needed. Work processes are constantly changing and the systems that are used in the processes must be flexible and cooperate with the user in order to avoid work arounds. To avoid time consuming work arounds, the participants in case study 2 asked for a flexible CMMS that is able to adapt to the current situation. This would not only save time but also reduce frustration in the work process. This pertains to context awareness within the CMMS. The operators need to receive the right information at the right time in the right context.

Thirdly, manual data input and transfer was a bottleneck in the system. To solve this, the participants in the case study 2 asked for a system that simplified common actions. This pertains to the user interface and the need for interface improvements.

Fourthly, the CMMS was not perceived as user friendly or intuitive due to complexity of the required actions and the fact that relevant guidance was lacking. The participants in case study 2 expressed the need for an intuitive interface and this pertains to interface improvements. Many problems could be solved by an intuitive interface, reducing user error.

Fifthly, lack of guidance pertains to the CMMS's lack of necessary information when operating the system. To combat this, the participants of case study 2 expressed the need for usability solutions that reflect common end-user solutions found in the marketplace today. And on top of that each action in the system should be confirmed by a sound or change in the interface giving relevant feedback on what action was taken so that the user knows exactly what has been done and what the next action should be, reducing incorrect actions. The job characteristic model lists relevant feedback as an important influence on the engagement of the employee [10]; by using positive and explanatory feedback the system can motivate the user to perform even better the next time data is fed into the system. This will increase intrinsic motivation [8][27] instead of just telling what has been done wrong and thereby only using extrinsic motivation to get the user to do the task correctly next time. Finally, the users need to know where they are in the system and where they are in the work process in relation to other tasks being completed. In this way they could plan their work more effectively and be able to quickly come up to speed.

Also, the cooperation between different teams would improve if other teams work actions are visualized. This is an important, if not necessary, part lacking in the CMMS. Being a part of that bigger picture and work together with different skills towards a common goal will increase the motivation and the sense of pride

of the work. It can also reduce unhealthy competition between groups within a company. The job characteristic model suggests that all different skills required for a task should be added to the task. This can encourage collaborations between teams [8]. The job characteristic model also describes the importance of task significance and task identity [8] that can be emphasized by the visualization of the whole work process. The system should be aware of the user context and able to visualize the context and the significance of the task. The users need to know the context so that they can be more effective in their work, as well as, reduce the chance for mental slips and mistakes caused by loss of context.

Sixthly, even though the CMMS is a powerful tool it was often only used in limited degrees for making strategic business decisions. The reason for this was the fact that the system was difficult to understand and use. Also, there was little transparency to the users what the result of the user actions actually could or should be used for in the complexity of the CMMS. The role of the user during each action should be known by the CMMS in order for the CMMS to be able to provide the purpose of what to be entered in the system. This would result in more accurate and relevant entered data. This pertains to the need of context awareness in the system. The users should be able to quickly orientate themselves by a glance or through a quick action request.

Lastly, the general lack of competence amongst the system users is a great concern. As noted in earlier studies, se [1], it is not unlikely that finances are expended in the implementation of the hardware and software, while training is not given the same level of importance. The participants in the case study described the CMMS training as an online self-study. Although this solution was most likely cheaper than a regular, hands on course with teachers and the actual equipment and system, relevant training is needed in order to gain user acceptance of a system.. To gain user acceptance, the purpose of why the new system is introduced must be thoroughly known of task significance and task identity are important for the user motivation [10]. If the training is a positive experience and performed in near time to when the user will start using the system, acceptance is more likely to occur. Therefore training should be considered as an important part of the system implementation.

The results from both case studies show that improvements need to be made within the area of user training, interface improvements and context awareness. The first, user training, shows that the case study participants did not feel that they were given the proper tools to work with the CMMS. The second, interface improvements, shows that there is a need for user-friendly design in many CMMS. As Labib described in [15], the systems seem to be rather accounting and/or IT oriented than engineering-based. The third area, context awareness, shows that the case study participants are need of a system that understands the complexity of the task that is about to be executed and the role of the current user of the system. The CMMS needs to have a context awareness solution and be able to sense what the user needs to accomplish a specified task [2][4][23]. Table 3 illustrates the relationship between the three identified core areas for the CMMS usability issues and the two case studies.

Table 3. Relation between core areas and case study results

Case study 1	Case study 2	Core area
Lack of compatibility	Interface should reflect end-customer solutions	Interface improvements
Data Multiplicity	Flexibility	Context awareness
Manual data input and transfer	Simplify common actions	Interface Improvements
Not user friendly	Intuitive interface is needed	Interface Improvements
Lack of guidance	Informative feedback is needed	Context awareness
	Orientation within the system is needed	
	Interface should reflect end-customer solutions	
Limited strategic use of CMMS	Allow for simple modifications	Context awareness
	Orientation within the system is needed	
Low competence	Better training needed	Better training

The need for user interface improvements points to the need for increased usability. As of today the CMMS interface is not intuitive and in some situations it can be rather difficult to use. There are EU directives [6] on how “useable” software programs should be and the main point of this directive is that the user should know what to do in each situation. This was not the case for the CMMSs studied. The greatest problem is not that the level of usability is low but that the users were prone to conduct errors because of the low level of usability. On top of that the users in several cases chose not to even use the CMMS but “push” the task of entering in, reading and using the data upon someone else. Design recommendations for complex systems are clear in that the user should see the system as enjoyable to use and not a threat [19].

Due to the complexity of the maintenance systems found in large industries, not just a better interface is necessary. Therefore it was stated by the users that they would like to work with a system that was able to know who was using the system and allow them to work with task relevant actions. Context awareness is necessary in assisting users in very large and complex systems like the CMMS of the companies that have been analysed in the studies.

Depending on who is using the system, different techniques might be used to raise the user motivation to add good quality input data. A typical user in this paper refers to e.g. a technician that is interacting with a CMMS and manually adding e.g. work order data, which will be used for further analysis and maintenance, to the system. However, usability aspects and context awareness are important for all interfaces, regarding if the user is a human or another system in the front or back end of a process.

5.1 Limitations

Several different CMMS manufacturers were used in the study. Since the user interfaces of these CMMS differ, there could be a conflict in the responses given by the case study participants. Although, the reported responses in these case studies were the general consensus of the personnel and the reported results congruent.

6. CONCLUSIONS

The conclusions of this paper are based upon how the CMMS can be improved. To improve the CMMS it was found that (i) interface improvements are necessary for the users to use the system more effectively, (ii) a greater focus needs to be placed on the training of the personnel and finally, (iii) the CMMS needs to have an intelligence that understands the context of the user situation and be able to sense what the user needs to accomplish a specified task.

The user aspects must be considered in relation to the current user. Different users of a CMMS may have varied expectations and needs and it is important to make the interface flexible and able to adapt to the user context. By implementing context awareness into a CMMS users are presented with information and allowed actions that are relevant to the goals they are to perform. Since the individual becomes an integral part in the maintenance process, with context adaption they are apt to a higher level of motivation and would be more likely to use the system to a higher degree.

7. FUTURE WORK

The interface improvements and training issues can be solved with an increased focus in those areas by the system designers and the businesses that implement the CMMS. It has been shown with today's end-users solutions, in both mobile and stationary solutions, that these issues can be reduced to a minimum. The area of context awareness is rather young and much research needs to be done so that industrial solutions reduce the risk for human error.

Usability aspects and context awareness are important for all interfaces, regarding if the user is a human or another system in the front or back end of a process. A subject for future work is to find a model or process describing of how a system interface can adapt to the user context in order to get good quality eMaintenance data.

8. ACKNOWLEDGMENTS

The authors would like to acknowledge Saab, Vinnova NFFP5, and F21 Wing Norrbotten for valuable comments and support in this study.

9. REFERENCES

- [1] Aljumaili, M., Tretten, P., Karim, R. and Kumar, U. (2012). "Study of Aspects of Data Quality in eMaintenance", Accepted to International Journal of COMADEM.
- [2] Baldauf M., Dustdar S. and Rosenberg F. (2007). "A survey on context-aware systems", *International Journal of Ad Hoc and Ubiquitous Computing*, 2(4), pp. 263-277.
- [3] Candell, O., Karim, R. and Parida, A. (2011). "Development of information system for e-maintenance solutions within the aerospace industry", *International Journal of Performability Engineering*, 7(6), pp. 583-592.
- [4] Dey, A.K. and Abowd, G.D. (2000). "Towards a better understanding of context and context-awareness", *In Proceedings of the Workshop on the What, Who, Where, When and How of Context-Awareness*, ACM Press, New York.
- [5] Dix, A., Finlay, J., Abowd, G. D., and Beale, R. (2004) "Human-computer interaction", third edition. Pearson Education Limited, Essex, UK.
- [6] EC Directive 90/270/EEC (1990). *Council directive on the minimum safety and health requirements for work with display screen equipment*, Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1990L0270:20070627:EN:PDF>, accessed: 8 August 2012.
- [7] Galar, D., Stenstrom, C., Parida, A., Kumar, U. and Berges, L. (2011). "Human factor in maintenance performance measurement", *In Proceedings of IEEM, IEEE*.
- [8] Gagné, M. and Deci, E. L. (2005). "Self-determination theory and work motivation", *Journal of Organizational Behavior*, 26, pp. 331-362.
- [9] Gustavsen, R.M. (2002). 'Condor – an application framework for mobility-based context-aware applications', *In Proceedings of the Workshop on Concepts and Models for Ubiquitous Computing*, Goeteborg, Sweden.
- [10] Hackman, J. R. and Oldham, G. R. (1976). "Motivation through the design of work: test of a theory", *Organizational Behavior and Human Performance*, 16(2), pp. 250-279.
- [11] Hofer, T., Schwinger, W., Pichler, M., Leonhartsberger, G. and Altmann, J. (2002). 'Context-awareness on mobile devices – the hydrogen approach', *In Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, pp.292-302.
- [12] ISO 9241-11, (1998). *Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) – Part 11: Guidance on Usability*, ISO.
- [13] Karim, R., Candell, O. and Söderholm, P. (2009). "E-maintenance and information logistics: aspects of content format", *Journal of Quality in Maintenance Engineering*, 15(3), pp. 308-324.
- [14] Karim, R. (2009). A service-oriented approach to e-maintenance of complex technical systems, Doctoral thesis, Luleå University of Technology.
- [15] Labib, A. W. (2004) "A decision analysis model for maintenance policy selection using a CMMS", *Journal of Quality in Maintenance Engineering*, 10(3), pp. 191-202.
- [16] Markeset, T. and Kumar, U. (2003). "Design and development of product support and maintenance concepts for industrial systems." *Journal of Quality in Maintenance Engineering*, 9(4), pp. 376-392.
- [17] Marquez, A.C. and Gupta, J.N.D. (2006). "Contemporary maintenance management: Process, framework and supporting pillars", *Omega*, 34(3), pp. 313-326.
- [18] Nelson, N. C. (2006). *Give Employees a Sense of Ownership*. Available at: <http://www.noellenelson.com/docs/American%20Venture%20Magazine%20-%202010-3-06.pdf>, accessed: 31 July 2012.

- [19] Norman, D. (1988). "The Design of Everyday Things", New York, Doubleday.
- [20] Prekop, P., and Burnett, M. (2003). "Activities, context and ubiquitous computing". *Computer Communications*, 26(11), pp. 1168-1176.
- [21] Reason, J. (1990). "Human Error", Cambridge, UK, Cambridge University Press.
- [22] Skipper blogspot (2012). *Interview with C PRIO Thomas Engevall regarding the introduction of PRIO in the Swedish Armed Forces*, Available at: <http://navyskipper.blogspot.se/2012/06/exklusiv-intervju-med-c-prio-thomas.html>, accessed on 1 August 2012.
- [23] Tanter, É., Gybels, K., Denker, M. and Bergel, A. (2006). "Context-aware aspects", Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 4089 LNCS, pp. 227-242.
- [24] Tretten, P., Karim, R. and Kumar, U. (2011) "Usability-based eMaintenance for effective performance measurement", *In Proceedings of MPMM*.
- [25] Quesenbery, W. (2001). "What Does Usability Mean: Looking Beyond "Ease of Use"", *Society Technical Communications Annual Conference*, 2001.
- [26] Wandt, K., Karim, R. and Galar, D. (2012). "Context adapted prognostics and diagnostics", *In Proceedings of CM2012/MFPT2012*.
- [27] Zichermann, G. and Cunningham, C. (2011). *Gamification by design*, O'Reilly Media, Inc., Sebastopol, CA.

Reliability Predictions based on Breakdown Data with Respect to Aging Conditions

Jakob Krause
 Member of Academic Staff
 Nöthnitzer Str. 46 Room 1076
 01187 Dresden - Germany
 +49 351 463-38529

Klaus Kabitzsch
 Full Professor
 Nöthnitzer Str. 46 Room 1074
 01187 Dresden - Germany
 +49 351 463-38289

jakob.krause@tu-dresden.de

klaus.kabitzsch@tu-dresden.de

ABSTRACT

In the area of eMaintenance the prediction of machine breakdowns is important in order to schedule maintenance actions efficiently. One part often missing in common predictive maintenance solutions is the consideration of the fact that machines may deteriorate non-uniformly, depending on aging conditions. In the paper *A generic Approach for Reliability Predictions considering non-uniformly Deterioration Behaviour* (2012) methods were presented showing how those aging conditions can be identified and used in order to improve the quality of reliability predictions. These Methods were based on condition monitoring data. In reality, condition monitoring data is not always available. In those cases reliability predictions are based exclusively on breakdown data not mentioning aging conditions. In this paper a concept is shown how reliability predictions can be made based on breakdown data with respect to aging conditions. Consequently, the accuracy of reliability predictions and maintenance actions can be improved.

Keywords

reliability predictions, aging conditions, breakdown data, aging contexts, survival analysis.

1. INTRODUCTION

During their whole lifetime machines are prone to aging processes. This leads to breakdowns and failures. Assuming machines (or machine components) are repairable, maintenance actions are executed which transfer them back into a healthy state. [4]

Thereby the challenge is to find the right point in time for maintenance actions. Predictive maintenance offers the possibility to predict the time of a break down and to schedule a maintenance action right before such an event occurs. Thereby, usually the health state of a machine is approximated based on condition monitoring data which indicates the aging process. Methods which are applied in those cases are methods of feature extraction (in the time- and frequency domain), pattern recognition, image recognition, time series analysis, ARIMA-Modeling, proportional hazard modelling and Hidden Markov modelling. An overview can be found in [1]. A problem is faced, in case no condition monitoring data is available.

In those cases conventionally maintenance actions are scheduled based exclusively on breakdown data. This approach is called survival analysis. The results of this analysis are statistical parameters like the MTBF which defines the time between two failures/break downs of a machine. [5]

Unfortunately, survival analysis has an important disadvantage which is the non-consideration of aging conditions (e.g. the drilling of a drilling machine in rock formations with different degrees of hardness).

As shown in [2] and [3] aging conditions influence the speed of aging of machines and therefore decrease the usability of statistical parameters. In order to determine accurate reliability predictions aging conditions need to be mentioned.

The objective of this paper is to show a method for reliability predictions which is based on breakdown data and aging conditions while condition monitoring data is not mentioned.

2. Definition of initial position and objective

In the focus of this method is a machine. This machine is prone to a continuously aging process. Consequently, the machine starts in an healthy state at time index $k=0$ and deteriorates until it breaks down at a time index $k=N_b$ with $k=0, 1, \dots, N_b$.

After a break down the machine is maintained and consequently transferred into a healthy state. Consequently, the machine goes through several operation cycles b with $b=0, 1, \dots, B$. Thereby, an operation cycle is defined as the interval between the time when a machine has been maintained until the time of its break down.

while k corresponds to a discrete time index the actual time t can be determined by equation 1 using the sampling interval T which defines the time between two samplings k and $k+1$.

$$t = k \cdot T \tag{1}$$

Figure 1 visualizes the operation cycles (Each reaching from the healthy state until a breakdown) and shows the introduced values so far.

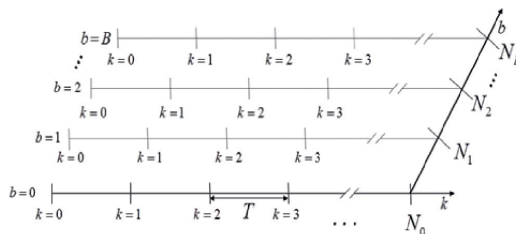


Figure 1: Operation cycles of a machine each reaching from the healthy state to a break down

As already mentioned (see also [2] and [3]) machines often do not age uniformly over time. Instead the speed of aging is depending on external influences. Those influences are here called contexts $C = \{c_i\}$ with $i=0,1,\dots,I$.

A context is here defined as a unique influence on the aging process of a machine. This means that a context cannot be divided into sub-contexts which have different influences on a machine. More information on contexts can be found in [3].

Furthermore, it is supposed that all contexts can be mapped to the indexes k_b of each operation cycle b . Figure 2 shows an example for an distribution of three contexts (c_1, c_2 and c_3).

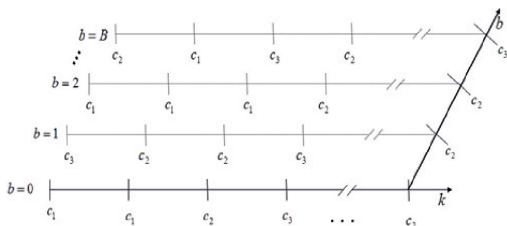


Figure 2: Example for a distribution of contexts

The objective of the proposed method is to determine the influence of contexts on the aging process in order to predict the time of break down for a given machine and given contexts.

3. Description of the Method

In this chapter the new method is described. As this is done in a very abstract manner in the following chapter a concrete example is shown. The here proposed method consists of two phases: a *training phase* and a *prediction phase*. Both phases consist of several sub-steps which are shown in Figure 3.

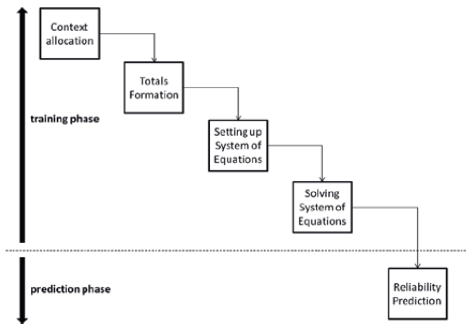


Figure 3: Steps for crating reliability predictions based on contexts and breakdown data.

The objective of the training phase is to determine a number for each context which defines the speed of aging under the influence of the contexts. In the following, the steps of the training phase are described in detail.

Context allocation: For each operation cycle b all influencing contexts c_i have to be determined. The result is a sequence of

contexts (equation 2) for each operation cycle (like shown in Figure 2).

$$(\dot{c}_{k,b}) = (\dot{c}_{0,b}, \dot{c}_{1,b}, \dots, \dot{c}_{n,b}), \dot{c}_{k,b} \in \{C\} \quad (2)$$

Totals formation: In this step for each operation cycle separately, the number of occurrence of each context is counted. These numbers are provided by the recursive function S (Equation 3) which has to be executed for each operation cycle and each context.

$$S(i, b, N_b) = \begin{cases} S(i, b, N_b - 1) + 1 & | c_i = \dot{c}_{N_b, b} \\ S(i, b, N_b - 1) & | c_i \neq \dot{c}_{N_b, b} \\ 0 & | N_b < 0 \end{cases} \quad (3)$$

Setting up system of equations: Based on the total numbers of each context and for each operation cycle a system of equations is set up (Equation 4). The vector b is filled with “ones” representing the healthy state of a machine. The state at the time of a breakdown/failure is indicated by the number “zero”. All other states between are mapped to a number in the interval $[1,0]$.

$$M \cdot \underline{x} = \underline{b}$$

$$\begin{cases} N(c_0, 0) & N(c_1, 0) & \dots & N(c_I, 0) \\ N(c_0, 1) & N(c_1, 1) & \dots & N(c_I, 1) \\ \dots & \dots & \dots & \dots \\ N(c_0, B) & N(c_1, B) & \dots & N(c_I, B) \end{cases} \begin{cases} x_0 \\ x_1 \\ \dots \\ x_I \end{cases} = \begin{cases} 1 \\ 1 \\ \dots \\ 1 \end{cases} \quad (4)$$

Approximation a solution: Usually the matrix M is over determined. That means there are more operation cycles than different contexts. For that reason, a solution for Equation 4 can only be approximated. This is done by mean of the *least squares method* (Equation 5).

$$\underline{x} = (M^T \cdot M)^{-1} \cdot M \cdot \underline{b} \quad (5)$$

The result is a number x_i with $x_i \in [1,0]$ for each context c_i . It indicates the influence of this context on the speed of aging whereas higher numbers indicate higher aging speeds.

After solving the system of equations the activities in the training phase are finished. In the following the prediction phase is described.

As precondition for the prediction the future occurrence of contexts must be known. This tuple is defined in equation 2 with $b=b_{future}$. Next, starting at $k=0$, k is iterative increased $k=k+1$ and for each k of the tuple the context c_i with its wight x_i is identified. Furthermore, the numbers x_i are summed up and subtracted from the number 1.0 which indicates the healthy state (equation 6).

$$d = 1 - \sum_k x_i \quad (6)$$

A breakdown is predicted at the time (k) when d reaches zero which is the break down state.

4. Example

In this example it is supposed there is a machine for which the time of breakdown has to be predicted. The aging process on this machine is determined by two contexts (c_1 and c_2).

The first step of the **training phase** is the **context allocation**. The result of this step is shown in Table 1.

Table 1: Example for the allocation of contexts to operation cycles

Operation cycle b	Sequence of contexts for ($k=0,1,\dots,N_b$)	Number of elements N_b
0	$c_1, c_1, c_1, c_2, c_2, c_2, c_1$	7
1	$c_1, c_1, c_2, c_2, c_2, c_2$	6
2	$c_1, c_1, c_2, c_2, c_2, c_1, c_1$	7
3	c_2, c_2, c_2, c_2, c_2	5
4	$c_1, c_1, c_1, c_1, c_1, c_2, c_1, c_1, c_1$	9
5	$c_2, c_2, c_1, c_1, c_1, c_1, c_1$	8

On the left side of the table the index numbers of the operation cycles are shown. In this example the reach from $b=0$ to $b=5$. On the right side the sequence of contexts until a breakdown of the concerned machine for each operation cycle are shown.

Secondly, the step of **totals formation** is executed. Consequently, by using equation 3 the number of occurrences of the contexts c_1 and c_2 for each operation cycle is computed. The result is shown in Table 2.

Table 2: Result of the totals formation

Operation cycle b	Occurrence of context c_1 $S(1, b, N_b)$	Occurrence of context c_2 $S(2, b, N_b)$
0	4	3
1	2	4
2	4	3
3	0	5
4	8	1
5	6	2

Thirdly, the step of **setting up system of equations** is executed. Therefore, the values from Table 2 are used in equation 4 which is shown in equation 7.

$$\begin{pmatrix} 4 & 3 \\ 2 & 4 \\ 4 & 3 \\ 0 & 5 \\ 8 & 1 \\ 6 & 2 \end{pmatrix} \cdot \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{Bmatrix} \tag{7}$$

In the last step of the training phase (**approximation a solution**) the system of equations shown in equation 7 is solved by mean of equation 5. The result is shown in equation 8.

$$\begin{aligned} x_1 &= 0,1 \\ x_2 &= 0,2 \end{aligned} \tag{8}$$

Thereby, x_j defines the speed of aging for c_j and x_2 for c_2 .

Next the **prediction phase** starts. In order to make a **reliability prediction** the future occurrence of contexts must be given. The future occurrence of contexts for this example is shown in Table 3.

Table 3: Future occurrence of contexts

$c_2, c_1, c_1, c_1, c_1, c_1, c_1, c_2, c_2$

The prediction is made using x_1 and x_2 (equation 8) and the future occurrence of contexts (Table 3) in equation 6. The prediction is shown in Table 4.

Table 4: Prediction of break down

k	context	x_j	$\sum x_i$	d
0	c_2	0,2	0,2	0,8
1	c_1	0,1	0,3	0,7
2	c_1	0,1	0,4	0,6
3	c_1	0,1	0,5	0,5
4	c_1	0,1	0,6	0,4
5	c_1	0,1	0,7	0,3
6	c_1	0,1	0,8	0,2
7	c_1	0,1	0,9	0,1
8	c_1	0,1	1,0	0,0
9	c_2	0,2		

The result of the prediction is, that for the given sequence of contexts at time $t=k*T$ with $k=8$ the health state ($d=0$) of the machine reaches a critical level which makes a breakdown most likely. Thus, a maintenance action should be executed before that time.

The MTBF for this example is determined in equation 9.

$$MTBF = T * \frac{\sum N_b}{B} = \frac{7+6+7+5+9+5}{6} = 6,5 * T \quad (9)$$

Supposing that the contexts have the above defined influence on the machine, a reliability prediction based on the MTBF is too early (at time $k=6,5$ compared to 8). Thus, a maintenance program based on the MTBF would waste on-time of the machine.

5. Conclusion

The method presented in this paper shows how reliability predictions based on breakdown data and aging conditions without mentioning condition monitoring data can be made. Furthermore, through the consideration of aging conditions the accuracy of reliability predictions is higher compared to those based on statistical parameters like the MTBF. Thereby, the greater the influence of aging conditions the more accurate the reliability prediction compared to the MTBF is. The whole method can easily be implemented in a software module for automatic generation of predictions. In this way maintenance actions can be scheduled more accurately which saves money and reduces down times.

Nevertheless, the here proposed method has been only tested on simple examples so far. Furthermore, the influence of contexts is only modelled linear.

Therefore, in a next step the method has to be tested on more complex use cases.

6. ACKNOWLEDGMENTS

Our thanks to sponsors of eMaintenance2012 for their intellectual and financial support.

7. REFERENCES

- [1] Andrew K.S. Jardine, Daming Lin, and Dragan Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7):1483 { 1510, 2006.
- [2] Krause, J and Kabitzsch, K. A generic approach for reliability predictions. *International Journal of Performability Engineering*, 7:553{562, 2011.
- [3] Krause J and Kabitzsch K 2010 *Context-sensitive Predictions of Machine Reliability* (Lulea – Sweden: The 1st international workshop and congress on eMaintenance)
- [4] Mobley R. Keith, *An Introduction to Predictive Maintenance – Second Edition*, Elsevier Science, 2002.
- [5] Zhanshan Ma, Survival, A.W.K., *Survival Analysis Approach to Reliability, Survivability and Prognostics and Health Management (PHM), Aerospace Conference, 2008 IEEE*, vol., no., pp.1-20, 1-8 March 2008
doi: 10.1109/AERO.2008.4526634

Condition Monitoring of Oil and Gas Subsea Electrical Equipment – Case study Åsgard Subsea Compression Project

Hermanto Ang, M.Sc.
University of Stavanger
N-4036 Stavanger, Norway
+47 4512 4861
hermanto_ang_w@yahoo.com

Tore Markeset, PhD
University of Stavanger
N-4036 Stavanger, Norway
+47 5183 2191
tore.markeset@uis.no

Tor Ole Bang-Steinsvik
ABB AS
Ole Deviks Vei 10, NO-0666, Oslo
+47 2287 2058
Tor-ole.bang-steinsvik@no.abb.com

ABSTRACT

The paper addresses the overall picture of condition monitoring in the case study of the Åsgard subsea compression project, related subsea electrical power and distribution systems, as well as high power rating subsea process systems. This includes: identification of failure modes in this equipment, suggestions for suitable condition monitoring techniques, data networks and integration alternatives to existing topside infrastructure and also suggestions for cost-effective maintenance management.

Keywords

Oil and gas; Åsgard subsea compression system; Condition monitoring; Subsea electrical equipment; Failure modes; Failure symptoms; Data network infrastructure; Data network integration; Maintenance strategy.

1. INTRODUCTION

Offshore oil and gas field development is shifting to the usage of subsea systems in situations where the field is located either in deepwater area or close to an existing field and can be connected to existing production systems. Subsea field development is trending towards subsea processing systems: the oil and gas processing equipment normally installed in the surface facility will be placed on the seabed closer to the well. This strategy is

expected to increase the reservoir recovery rate, especially in marginal fields and also to reduce total capital expenditure by eliminating the topside processing cost at the same time. Meanwhile, the processing equipment normally requires a high electrical power supply, for which there are dedicated subsea electrical power and distribution units. Once the equipment is installed on the seabed, the cost of inspection and maintenance of this equipment increases exponentially. However, condition monitoring may enable the operators to predict the condition degradation and to plan a cost-efficient and effective maintenance strategy, with the maintenance actions being conducted based on the actual health of the equipment. The identification of subsea equipment failure modes is the key to efficient and effective condition monitoring. Potential benefits of condition monitoring include [1]: maintenance cost savings, reduced repair time and machine or plant downtime, better utilization of machinery, reduced personnel risks, improved profitability. Subsea processing has emerged as an interesting method of field development, with the purpose of increasing the recovery rate of oil and gas production from the reservoir. According to [2], subsea processing can be defined as any active treatment of oil and gas on the seabed. There are several alternative means of treating oil and gas on the seabed, each with their own technological maturity, as shown in Table 1.

Table 1. Subsea Processing System Technological Maturity.

Subsea Process	Status	Experience
Multiple boosting	Maturity technology	Several successful applications such as Draugen field NCS, King field GoM
Gas liquid separation and liquid boosting	High technical maturity technology	Subsea operational experience, for example Tordis field NCS
Bulk water separation	High technical maturity technology	Subsea operational experience, for example Troll Pilot project NCS
Complete water separation	Some further technology maturity required	Prototype installed in Marlim Field Brazil
Gas compression	Extensive further technology maturing required	Ongoing Ormen Lange subsea compression full scale testing in Kårstø K-Lab and Åsgard subsea compression project 2014

Several beneficial possibilities can be gained through the implementation of subsea processing systems, such as:

- Reduced total CAPEX by eliminating topside and pipeline cost
- Increased production and recovery rate of oil and gas reservoirs
- Enabling marginal field developments, especially in deep water fields and fields with long tie-back distance
- Extended production from existing fields

- Enabling tie-in of satellite field developments into the existing infrastructures.

The Åsgard oil and gas field is located in the Norwegian Sea, 200 kilometers offshore at a water depth of 240 – 300 meters. The field has been developed with subsea completed wells tied back to the Åsgard A production and storage vessel. In addition, a floating semi-submersible facility (Åsgard B) is used to process gas and condensate, and a vessel is employed for storage of condensate (Åsgard C). For more details see [3].

Analysis stated in [4] shows that the gas pressure will become too low to sustain their ability to lift the gas to Åsgard B platform towards the end of 2014. The Åsgard subsea compression project, which will increase the recovery by an estimated 28 billion m³ of gas and 2.2 billion m³ of condensate, will be the world's first subsea compression system when it is installed in 2014 in the Norwegian Sea.

The subsea compression station will be built using a modularized system to increase maintainability should there be a need to replace components or subsystems. The subsea compressor station with two identical compression trains contains the inlet and anti-surge cooler module, the export gas cooler module, the gas-liquid separator/scrubber module, the compressor module, the subsea power and control module, the liquid pump module, the train subsea control module, the compressor anti-surge control module, and the active magnetic bearing control module.

The subsea compression station, which comprises of a large power rating and high frequency compressors and pumps, will be powered up by a frequency converter located on the Åsgard A platform. The electrical power from the topside power distribution grid will be connected to a step-up transformer and frequency converters or VSDs (variable speed drive) for subsea compressor and pump, all located topside. A 47 km subsea power umbilical connects the topside equipment to the subsea step-down transformer before distribution to the subsea equipment.

From an electrical and control point of view, the Åsgard subsea compression project is challenging due to:

- High or even highest power rating for a step-out system with 15 MVA transmission power.
- The high speed subsea pumps and compressors need 200 Hz power supplied from the topside frequency converter.
- Highest supply frequency for the 47 km step-out cable connected to topside frequency converter output.
- Transmitting power at variable frequency to variable loads (pumps and compressors) through long subsea cables represents many challenges.
- The subsea compression station will be controlled using an all-electrical control system.

2. IDENTIFYING FAILURE MODES ON ÅSGARD SUBSEA COMPRESSION SYSTEM AND TECHNIQUES FOR MONITORING SYMPTOMS DEVELOPING TOWARDS FAILURES

In this section, failure modes of some chosen Åsgard subsea electrical power and distribution units, as well as of high power rating subsea processing equipment, will be identified together with the means of detecting these failure modes. The definition of failure modes, according to Norsok Z-008, is the effect by which a failure is observed on the failed item. Mapping of the failure modes is presented in Table 2 for subsea electrical equipment and in Table 3 for subsea processing equipment.

3. SUBSEA CONDITION MONITORING DATA NETWORKS AND INTEGRATION

ALTERNATIVES TO EXISTING TOPSIDE INFRASTRUCTURE

3.1 The Challenges in Network Communication for Subsea Development Integrated to Topside Systems

Åsgard subsea compression module, which will be integrated with both Åsgard A FPSO and Åsgard B semisubmersible, brings together additional challenges for SAS (Safety and Automation System) and CM (condition monitoring) integration network besides their own natural challenges from its subsea deployment. Some challenges from Åsgard subsea compression module development include:

- Integration of subsea control system SAS and CM network into two different surface production facilities, with equipment being installed in Åsgard A and controller in Åsgard B, which will increase the complexity level.
- Segregation of CM network from PCS (process control system) and PSD (process shutdown) networks in order to overcome high bandwidth requirements which will introduce several new pieces of equipment to be installed such as routing switch, CM SEM (Subsea Electronic Module), fiber optic interface penetrators and thus will increase complexity.
- Integrated fiber optic subsea system for CM data transmission which will increase the financial investment for the oil company.

3.2 Important Features to be Implemented in The Infrastructures

According to ISO 13628-6 (ISO, 2006) concerning the Design and Operation of Subsea Production Control Systems, redundancy is needed to such a level that it can prevent or minimize the loss of subsea production due to a single component failure or common mode failure, and the level of redundancy throughout the system is influenced by the complexity and reliability. Analysis of the expected benefit from redundancy should be performed for all critical parts of the system. Based on the Åsgard Subsea Compression Project system specification requirements and philosophies, the control system will be completely redundant (A and B systems) including network redundancy (PCS, PSD or CM), network channel redundancy, communication redundancy and controller redundancy.

The other important features are to utilize high-bandwidth fiber optic systems as a communication backbone between subsea SCMs and topside monitoring units. The Åsgard subsea compression project utilizes 100 Mb TCP/IP communication protocol on separate fibers. Data from condition monitoring field sensors is transmitted to the SEM installed in the SCM via optical fiber. Then, communication between subsea sensors' monitoring systems and subsea production control systems, as well as subsea transmission systems to the surface production facility will utilize a fiber optic communication link. The whole system of fiber optic communication is known as IFOSS (Integrated Fiber Optic Subsea System).

Normal set-up for fiber optic communication consists of computer/control system connected to a 485/232 converter for serial communication and then to a fiber optic modem. The distribution fiber extends from the fiber optic modem to another

far-end receiving modem and back to the 485/232 converter. The new technical challenge is introduced when the fiber optic from the far-end receiving modem and converter is connected to the SCM. Terminal equipment such as transmitters or receivers has to be incorporated into the SCM, and subsea wet-mateable connectors have to be incorporated to allow standard umbilical terminations and connection to retrievable control modules.

3.3 Different Alternatives of Åsgard Subsea Compression Data Integration

Åsgard has two existing surface production facilities, Åsgard A FPSO and Åsgard B semisubmersible. According to the Åsgard Subsea Compression Project system specification requirements and philosophies, topside equipment related to Åsgard SCSt (subsea compression station) will be installed in Åsgard A and will be controlled from Åsgard B.

There are several alternatives to integrating the Åsgard subsea compression station to the existing Åsgard SAS network, as shown in Figures 1 to 4.

4. PLANNING COST-EFFICIENT MAINTENANCE STRATEGIES BASED ON CONDITION MONITORING DATA

Data from the condition monitoring system can be utilized to plan the maintenance strategy in the most effective schedule and necessities for known downtimes before the failure actually occurs. This maintenance strategy is more widely known as condition based maintenance. It has become more critical to implement condition based maintenance in the subsea oil and gas industry, since the cost incorporated in the maintenance activity is usually increased by additional factors, besides the cost of deferred production, which have to be considered such as cost of intervention vessel, cost of renting specialized subsea tools as well as ROV (remotely operated vehicle) tools, cost of transportation for spare parts or maintenance technician, cost of waiting on weather and also cost of regulatory safety check.

Knowledge from the Åsgard subsea compression condition monitoring system and expected failure distribution provides a very broad opportunity for the oil company to implement condition based maintenance strategy on the Åsgard subsea compression system. This paper suggests a condition based

maintenance flowchart, which gives an overview of the whole picture of effective maintenance strategy. See Figure 5.

There are several factors which have to be considered when planning and executing condition based maintenance strategy, such as equipment condition degradation rate, criticality of degradation, review of affected systems or subsystems. After having knowledge on the technical degradation and criticality, the oil company should develop and review maintenance alternatives, weather window constraints, current vessel assumptions the company has. Before conducting actual operation offshore, the oil company should develop IMR (inspection, maintenance and repair) tasks and procedures to be followed; ROV, vessel, and tooling requirements; as well as regulatory safety checks after IMR activities have been carried out.

5. REFERENCES

- [1] Markeset, T., (2011), Condition Monitoring and Management Lecture Notes, Norway: University of Stavanger.
- [2] Fantoft, R., (2005), Subsea Gas Compression - Challenges and Solutions, In: *The Proceedings of Offshore Technology Conference*, Houston 2-5th May 2005, OTC 17399.
- [3] Statoil, (2007), *Åsgard*. <Available at: <http://www.statoil.com/en/ouroperations/explorationprod/ncs/aasgard/pages/default.aspx>>
- [4] Halvorsen, G.R., (2011), Åsgard Subsea Compression Control System: Solutions and Challenges. In: *Offshore Instrumenting 2011*.
- [5] ABB Inc., (2007), *MCM800 Machinery Condition Monitoring: Addressing The Needs of Your Rotating Machinery*, USA: ABB.
- [6] Østergaard, I., Nysveen, A., Romanisko, T., (1999), Mecon: A High Voltage Subsea Connector, In: *The Proceedings of Offshore Technology Conference*, Houston, 3-6th May 1999, OTC 10948.
- [7] ABB AG, (2012), *Modular Switchgear Monitoring: SF6 Gas Monitoring for Gas Insulated Switchgear*, Germany: ABB.
- [8] Ang, H., (2012), Condition Monitoring of Oil and Gas Subsea Electrical Equipment – Case Study Åsgard Subsea Compression Project, MSc thesis, Norway: University of Stavanger.

Table 2. Mapping of Failure Modes and Monitoring Techniques for Subsea Electrical Equipment.

Equipment	Failure Identification		Condition Monitoring Technique
	Possible failure modes	Symptoms of failure	Field sensing devices
Subsea step down transformer	Tap changer wear	Overnumbering of contact operation	ABB TEC internal algorithm
	Excessive moisture content in oil	Increasing trend in moisture reading	Moisture in oil sensor
	Excessive hydrogen content in oil	Increasing trend in hydrogen reading	HYDRAN for hydrogen gas in oil sensor
	Earth fault	Significant increase in earth current	Earth fault current monitor
	Overheating	Increase in oil temperature, winding temperature, current	Oil temperature sensor, current transducer, ABB TEC internal algorithm
	Insulation damage	Excessive hydrogen and moisture content	Moisture in oil sensor, HYDRAN, ABB TEC
	Winding fault	Increase in leakage factors, winding temperature, winding current	ABB TEC internal algorithm, current transducer
Subsea electrical motor	Stator faults	Stator magneto motive force (MMF)	Current and voltage sensors
	Rotor faults	Unequal rotor bar currents	Rogowski coil measurement
	Insulation breakdown	Decreasing electrical resistance, moisture content in the windings	Megaohm resistor, moisture sensor
	Overheating	Increasing stator and rotor temperature	Stator winding temperature detector (RTD) and

Equipment	Failure Identification		Condition Monitoring Technique
	Possible failure modes	Symptoms of failure	Field sensing devices
Electrical penetrator and wet mateable connector	Eccentricity faults, consisting of: Mechanical imbalance, misalignment, bent shaft	Increasing machine vibration	lumped thermal parameter Current, voltage sensors and ABB Machsense
	Water intrusion	Dropping dielectric breakdown voltage	Withstand voltage meter
	Earth fault	Significant increase in earth current	Earth fault current monitor
	Insulation fault	Decreasing dielectric breakdown voltage	Withstand voltage meter
	Fault in electrical strength	Dropping dielectric breakdown voltage	Withstand voltage meter
	Interfacial breakdown between dielectric surfaces	Dropping dielectric breakdown voltage	Withstand voltage meter
Subsea power cable	Stretched and bent due to cable stress	Increasing temperature in the cable	Distributed thermal sensing (DTS)
	Wear and tear due to dynamic in the cable	Increasing temperature in the cable	Distributed thermal sensing (DTS)
Subsea switchgear	Dielectric breakdown	Decreasing gas pressure and density, increase in humidity and temperature	ABB MSM internal algorithm, moisture and temperature sensors
	Mechanical faults	Increasing vibration, contamination and moisture content in moving parts	Moisture sensor, vibration sensor
	Excessive temperature	Increasing temperature in switchgear conductors and gas insulation	ABB SensyCal FC400 IR, ABB MSM
	Arc faults	Oxidizing contact surface, rapid increase in temperature in contactor	ABB MSM internal algorithm
	Partial discharge	Excessive humidity and moisture content	Moisture sensor, withstand voltage meter

Table 3. Mapping of Failure Modes and Monitoring Techniques for Subsea Process Equipment.

Equipment	Failure Identification		Condition Monitoring Technique
	Possible failure modes	Symptoms of failure	Field sensing devices
Subsea gas compressor	Bearing failure	Increasing in compressor vibration, bearing temperature	Temperature sensors, ABB MCM800
	Shaft failure	Increasing compressor vibration, friction and wear	ABB MCM800
	Overheating	Increasing temperature in rotating parts, high compression ratio, high return gas temperature	Temperature and pressure sensors in compressor inlet and outlet
	Internal corrosion	Increasing vibration in compressor	ABB MCM800
	Clogged suction strainer	Increasing pressure differential in filter	Pressure differential sensor
	Loss of gas output	Dropping in discharge temperature, compressor efficiency	Temperature, pressure and flow sensors in compressor inlet and outlet
	Surge and cavitation	Low flow rate at compressor inlet, increasing compressor vibration	Pressure and flow sensors in compressor inlet and outlet
Subsea liquid pump	Drop in produced head	Decreasing reading of pump pressure sensor, increasing pump vibration	Pressure sensors, ABB MCM800
	Shaft failure	Increasing pump vibration	ABB MCM800
	Pump wear	Dropping in pump produced head and pump performance, increasing leakage path internal to the pump and friction	Pressure sensors, flowmeter, operating head
	Pump overheating	Increasing temperature in rotating parts	Temperature sensors
	Corrosion	Increasing pump vibration	ABB MCM800
	Loss of liquid output	Dropping in pump produced head, decreasing pump performance	Pressure sensor, flowrate, operating head
Control valve and electric valve actuator	Erosion in body and trim	Increasing noise and vibration in the valve	Vibration sensors
	Overheating	Increasing temperature of internal gear and motor	Embedded temperature sensors
	Cavitation	Increasing noise and vibration in the valve, high flow rate and low pressure	Vibration sensors, flowmeter, pressure sensor
	Drive signal failure	Decreasing reading in actuator input voltage and current	Voltage and current sensors
	Valve trim travel deviation	Insufficient torque, actuator position feedback deviation	Torque monitor, resolver feedback positioner
	Slow response time	Increasing time delay between control signal and valve position feedback	Internal valve self diagnostic timer
	Loss of communication	Insufficient feedback signal from valve actuator	Resolver feedback positioner

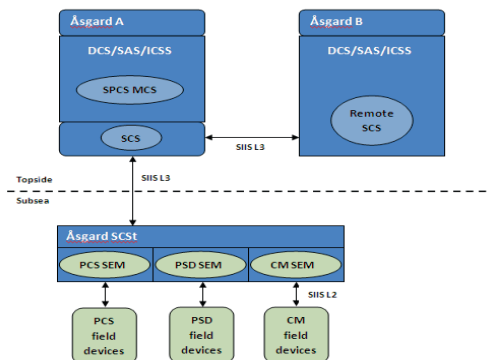


Figure 1: Åsgard A integration (Logic server in Åsgard A and controller in Åsgard B)

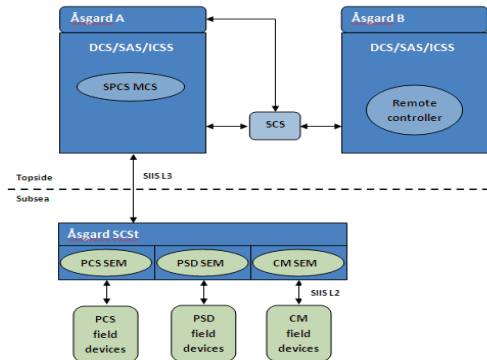


Figure 3: Standalone Åsgard SAS (New dedicated infrastructure for condition monitoring)

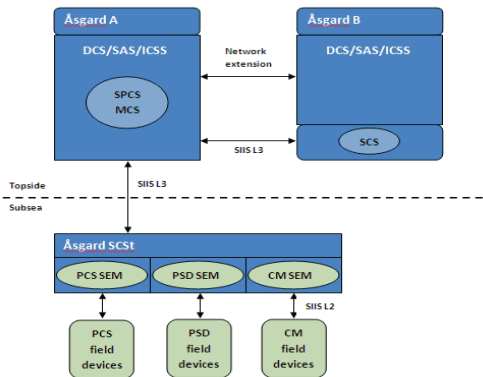


Figure 2: Åsgard B integration (Logic server and controller in Åsgard B but need to extend infrastructure to Åsgard A)

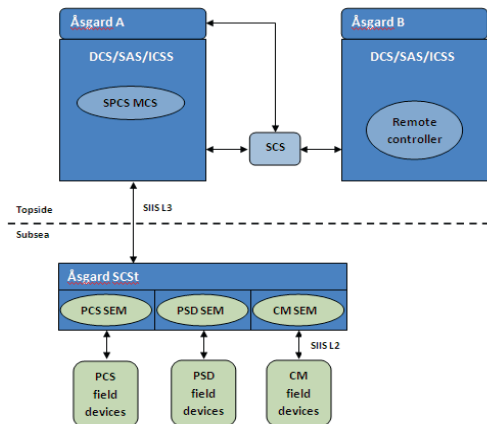


Figure 4: Standalone Åsgard SAS (New dedicated infrastructure for condition monitoring)

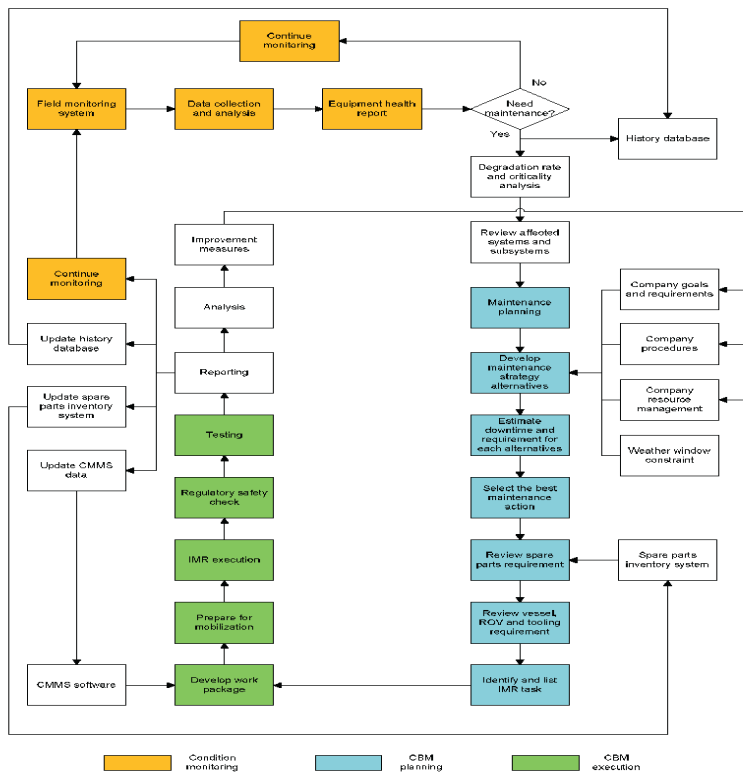


Figure 5: Flowchart of condition based maintenance in subsea oil and gas industry

A Study of Wireless Vibration Sensors for Monitoring Bearing Faults

Mustafa Aliwan
University of Huddersfield,
Queensgate, Huddersfield HD1
3DH, UK
00447412062616
A.Mustafa@hud.ac.uk

Fengshou Gu
University of Huddersfield,
Queensgate, Huddersfield HD1
3DH, UK
00441484] 472965
f.gu@hud.ac.uk

Andrew Ball
University of Huddersfield,
Queensgate, Huddersfield HD1
3DH, UK
00441484 473640
a.ball@hud.ac.uk

ABSTRACT

The reliability and sustainability of machine systems can be improved by condition monitoring techniques. Rolling bearings are among the most important and frequently encountered components in the vast majority of rotating machines including railway machines. Their deterioration are prominent for the overall machine performance.

The conventional method to monitor bearing is based on a wired accelerometer mounted to the bearing houses. However, in many harsh situations, it is difficult to use a wired accelerometer to collect data directly from the bearings. In recent years, the development in wireless communication technology makes it possible to measure vibration remotely at low initial cost for monitoring not only the hard situations but also with promising accuracrobblems with wiring related faults.

In this paper a wireless data acquisition system is examined based on monitoring bearing faults. Unlike many low bandwidth wireless systems, the wireless accelerometer node whi under investigation has multiple channels and is able to sample vibration data simultaneously at a rate as high as 10 kHz and hence it is possible to monitor bearing vibration in a high frequency band to obtain accurate diagnostic features. Common spectrum and envelop spectrum are used to analyse the vibration data which are collected from wired and wireless system on the outrace of a roller bearing. The results show that the data from wireless system allows the identification of bearing faults with a performance nearly as good as a wired system. However, a wider bandwidth of the wireless node may be appreciated for more accurate bearing detection.

Keywords

Wireless accelerometer node; bearing vibration; bearing fault detection.

1. INTRODUCTION

Rotating machinery is a vital part of every manufacturing process such as motors, gearboxes, bearing, pumps, and compressors. They are relied upon to operate efficiently to maintain a steady stream of production at maximum throughput. The lifespan of every piece of machinery is limited by the speed at which it is operating, the load to which it is subjected, the quality of its components, assembly and installation, its environment and the level of maintenance. In recent years this view has changed to include maintenance and the management of maintenance being an integral part of the operation, production and subsequently, profits of the company. This change has resulted in the

development of several philosophies that aim to establish the most appropriate maintenance strategy for defined failure modes. One of the most widely used techniques is condition based maintenance. This technique defines maintenance actions based upon the change within one of or group of defined monitoring parameters. By virtue of its direct relevance to the condition of rotating machinery and its diagnostic value, the principal parameter used in monitoring is vibration [1],[5].

The advantages of wireless technology are to facilitate applications that were in the past impractical, such as temporary installations for troubleshooting and remote monitoring. On the other hand the challenges to Wireless Vibration Monitoring are the unique demands on wireless devices, networks and associated components. High bandwidth is needed, due to the relatively large amounts of data that need to be sent over the wireless link. In addition, good dynamic range, low noise levels and higher-level processing capabilities, and the ability to capture data at the right time are also key requirements. Battery-powered devices that are required to provide on board power must satisfy customer demands for long service life. The devices and sensors, as well as the wireless network components, must also cope with conditions commonly found in the industrial environment [2].

Vibration is a crucial problem for accurate machining processes, especially with rotating machine processes which need accurate results for many different purposes. Therefore, any change of vibration of machine tool parts will have a great influence on surface quality of the bearing at rotating machine processes. Using advanced vibration sensor monitoring and automation technology is an effective way to improving the product quality of rotating machine. Condition monitoring is a main method for accurate and reliable data acquisition of vibration. Over the years, various methods have been proposed to achieve machine tool condition monitoring, and recently wireless sensor networks based on condition monitoring have become highly popular. Wireless sensor networks consists of many distributed sensors called nodes, these nodes are used to monitoring or detecting various kinds of changes in vibration, pressure, temperature or movement etc. This report presents a vibration monitoring system of bearing machine tool based on wireless sensor networks (WSNs). Some key issues of wireless sensor networks for vibration monitoring system of rotating machine are discussed. The wireless transmission of vibration signals is achieved by the combination with vibration sensor nodes and wireless network [3],[4].

2. ARCHITECTURE OF THE TEST RIG

The test rig that been chosen under this project is shown in Figure 2. It consist of five main parts; electrical induction motor, coupled, dynamic brake, Bearing and Shaft. The motor is connected to the brake by shafts, which are connected by three pairs of matched flexible couplings. Two bearing housings each contain one roller and one captive ball bearing supporting these shafts.

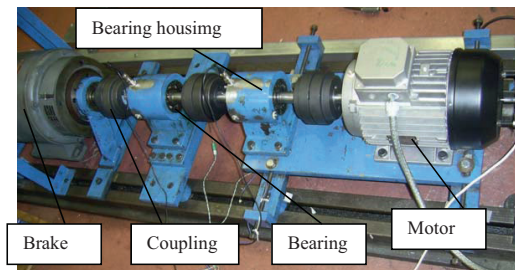


Figure 1. Architecture of the Test Rig

The rolling bearings (N406) are used under this test and its specification is shown in Table1. Rolling bearing consists of four main components: outer race, inner race, cage and rollers as shown Figure1.

Table1. Rolling Bearing (N406) specification

Elements	Measurement
Roller element Diameter	14 mm
Roller Number	9
Pitch Diameter	59 mm



Figure 2. Rolling Element Bearing.

It can be seen that from the table 2 the frequencies characteristic which is usually based to diagnose faults from this bearing in envelope spectrum analysis. This type of bearing can be separated

easily into different parts; various faults can be created conveniently for study.

Table2. Frequency Calculation Results for Bearing (406)

Defect position	Frequency (Hz) at 1460 rpm of the shaft
Inner race	135.1
Outer race	83.3
Rolling elements	48.3
Cage	9.3

2.1 Data Acquisition System

A data acquisition system (DAS) is a device designed to calculate and monitor parameters such as temperature, vibration, sound, etc. The DAS generally has two parts: hardware and software. It is mostly used to collect data and has some basic data analysis tools as well such as spectrum calculation for online data examination. This hardware is consisting of three parts. The first part is the sensors which are piezoelectric accelerometer, which has a frequency band up to 10 kHz, and the second one is a data acquisition card (DAC) that installed in PC computer and the amplifier, which is an important element in measuring vibration. The amplifier provides two advantages, first it increasing the vibration signal, which is often very weak, secondly, it isolates between the processing and display equipment and the vibration pick up. However, a charge amplifier is selected to produce a voltage which is directly proportional to the applied mechanical stress[10].

During the experimental work all the collected data was acquired using Global Sensor Technology YE6232B. This system has 16 channels each channel with a 16 bit analogue-digital converter with a maximum sampling frequency of 96 kHz per channel. The specification of this system can be found in table 3. The DSA YE6232B model is capable of analysing data in different ways such as time domain or frequency domain via a Fast Fourier Transform. It was only used for data acquisition and recording. Most of complicated signal processing techniques were carried out using MATLAB. During the data acquiring process the number sampling frequency were set to 12kHz respectively. For such setting the required time to collect this length of data is found to be 20 sec hence the frequency resolution is 0.4Hz.

A data acquisition system (DAS) is a device designed to calculate and monitor parameters such as temperature, vibration, sound, etc. DSA YE6232B model is capable of analyzing data in different ways such as time domain or frequency domain via a Fast Fourier Transform. Channels 3 and channel 4 are used from 16 channels which connected to the vibration sensors.

The DAS generally has two parts: hardware and software. It is mostly used to collect data and has some basic data analysis tools as well such as spectrum calculation for online data examination. The hardware consists of two parts: the vibration sensors and data acquisition card (DAC). The software is Matlab that is installed in PC computer.

Table 1. Data Acquisition System specification

Manufacturer	Global Sensor Technology YE6232B
Number of channels	16 channel (Selectable voltage/IEPE input)
Sampling rate (Max.)	96 KHZ
Input range	10V
Gain	Selectable (1, 10 or 100)
Filter	Anti-aliasing filter
Interface to host PC	USB 2.0

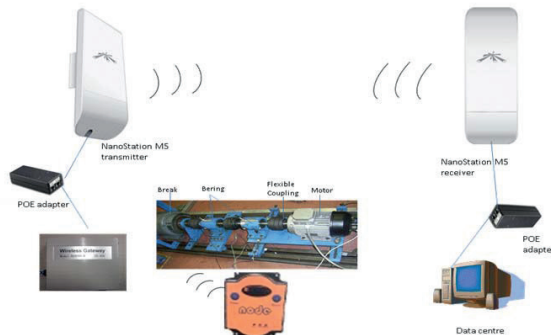


Figure3. The architecture of the monitoring system

2.2 Test Rig Operating Conditions

This test shows the methods and techniques of using wired and wireless system to monitor bearing condition. Two bearing conditions were considered, one is connected to the motor side, and the other is connected close to the generator. This experiment is implemented to detect the fault on the outer race of the bearing (see fig 6).

The experiment was conducted under a shaft rotational speed (SRS) of 1461.7 rpm or 24.36 Hz. This test was implemented twice to collect the data from a wired network and a wireless sensor network. Firstly, there are two channels of vibration signals collected from the accelerometers mounted on the bearing housing in a horizontal position with a load of 50 % and sampling rate of 12000. Secondly, the wireless sensor network test is implemented to make comparisons with the wired. This test is installed to monitor the vibration signal via the wireless network. This involves using two vibration sensors mounted horizontally to the bearing and connected to the sensor node with a sampling rate 10000Hz.

The performed test involves placing a scratch that is 100% of the bearing’s outer race width (seeded fault) see figure 6. Two bearing are installed on the rotary machine. The first bearing is healthy and is located close to the motor and the second bearing is faulty and connected close to generator.

2.3 Overview of Wireless System Architecture

The diagram below shows a complete system to meet the requirement for different measurements including vibration, temperature, stress, voltage, etc. A wireless network is usually composed of four of products:

- wireless sensor node(vibration)
- wireless gateways (base station) BS909N Main Controller (Base Station 406) connected with transmitter
- Control center PC connected with receiver
- Control Software

Sensor node is capable of detecting and collecting the raw data of vibration signals at test rig. These data are processed to amplify voltage, filter noises and operate A/D conversion. After that sensor node transmits these data to BS909N Main Controller via WSNs. BS909N Main Controller is a wireless gateway node, which is responsible for receiving the sensor data and retransmitting these data to the control center PC via Nano Station M5 5GH Indoor/outdoor MIMO 16dBi CPE Receiver/Transmitter module. The control center PC is responsible for processing, displaying and storing the sensor data.

2.4 Sensor Node

Sensor node consists of a micro-controller chip MCI3213 and vibration measurement sensor. The MC13213 is Free scale Semiconductor’s second-generation ZigBee platform incorporates an 8-bit microcontroller and 2.4 GHz radio frequency transceiver into a single package. In addition it contains a low noise amplifier, 1mW output power, with internal voltage controlled oscillator, on-board power supply regulation, integrated transmit/receive switch, and full spread-spectrum encoding and decoding unit. Its main features include fully compliant 802.15.4 Standard transceiver which supports 250 kbps O-QPSK data in 5.0 MHz wideband channels, and operates on the license-free ISM frequency band in the 2.4 GHz, and ZigBee-compliant network stack, etc. In machining, real-time vibration data will be collected through vibration measurement sensor (accelerometer). Vibration sensor is an accelerometer (CA-YD-109B) with a measurement range of 0.2-1.5kHz and sensitivity of 289 pc/ms-2.

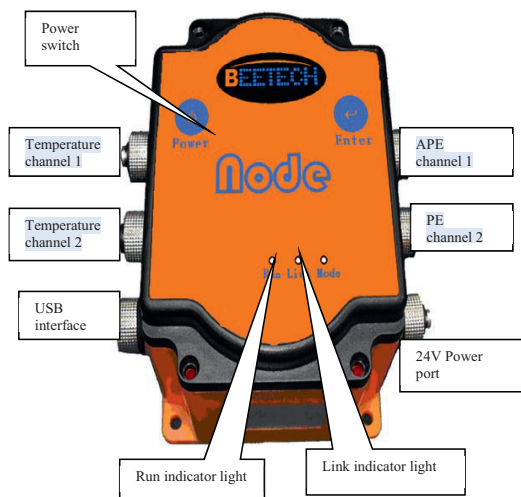


Figure4. An image of sensor node

2.5 Gateway

The main controller BS909N is a gateway developed by Beijing BeeTech Inc as it is shown in Fig.3. It complies with TCP/IP and ZigBee protocols. It is capable to work in a TCP/IP complied office area. BeeTech's wireless sensors send the data to BS909N gateway via ZigBee network and BS909 gateway then sends the data collected by the sensors to the designated server data centre. BS909N gateway is designed with DC power supply which needs to plug in the power adapter to the "Power" input of BS909N gateway while using it. All data from every sensor node to coordinator can be transmitted via ZigBee network transmission protocol. In addition, coordinator is responsible for network setup, communication management, information store, and so on [6].



Figure5. An image of Gateway

2.6 Transmission Protocol

Zigbee is a standard for Wireless Personal Area Network (WPAN) applications that require low data rates and long battery life. Such as a wireless sensor network that can be used to acquire condition data which may last several seconds per day or so.

ZigBee is essentially for battery-powered applications where low data rate which is 250 K bits per second, low cost, and operate in 868 MHz, 915 MHz, and 2.4 GHz frequency bands. The ZigBee Alliance is an association of companies which develops standards and products for reliable, cost-effective, low power wireless networking. The ZigBee standard is built on top of the IEEE 802.15.4 standard. The 802.15.4 standard defines the physical and MAC (Medium Access Control) layers for low-rate wireless personal area networks. ZigBee standardizes both the network and the application layer. The network layer is in charge of organizing and providing routing over a network, specifying different network topologies such as star, cluster tree or mesh networks. The Application Layer provides a framework for distributed application development and communication [7].

2.7 Software Used

The server side software was developed by installing Bee-Data software on the PC. It creates a virtual communication port to the gateway and displays the incoming data packets in various formats. Bee-Data software should open and by selecting "IP address and Port Type which is "192.168.1.100: 15002" this should match the data centre server IP address.

3. RESULTS AND DISCUSSIONS

3.1 Wired System

The spectra were calculated using FFT algorithm and employing a Hanning window with 50% overlap processing. All figures of spectra are plotted in the frequency range of 3900-5800 HZ to filter the signal. The envelope analysis method is implemented on the defects in the bearing and therefore the fault frequencies of high amplitudes with their harmonics will appear in the envelope spectra. Since the highest fault frequency is (83.86 Hz) which is the outtrace fault, all envelope spectra plotted here are in the frequency range.

Root-Mean-Square (RMS) is used as an indicator of average amplitude level of vibration signals. As the energy within a signal is relative to the squared value of vibration amplitude, RMS can also be considered as a directory of vibration energy in this study.

On the other hand, it can be seen that from the Figure7 the blue line represent that the bearings normal condition means that the shaft is rotating at a constant speed, supported by healthy bearing. The first impression created by the waveform in Figure 7 is that the general vibration level of the signal is slightly higher than those of the normal condition with a RMS equivalent to 3.051.

The characteristic frequency of the outer race defect should be obvious in the spectrum. However, as Figure 8 shows in red line, these fault frequencies are masked by 1 x SRS and its harmonics because of their high amplitude. In addition, there is a significant decrease in the amplitude.

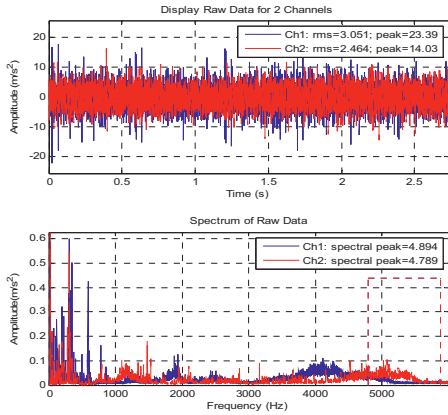


Figure6. Weird Raw Data and Spectrum for Raw Data

The vibration signal from the early stage of a defective bearing may be masked by machine noise which is making it difficult to detect the fault by spectrum analysis alone. The main advantage of envelope analysis is its ability to extract the periodic impacts and the modulated random noise from a deteriorating rolling bearing. This is even possible when the signal from the rolling bearing is relatively low in energy and ‘buried’ within other vibration from the machine.

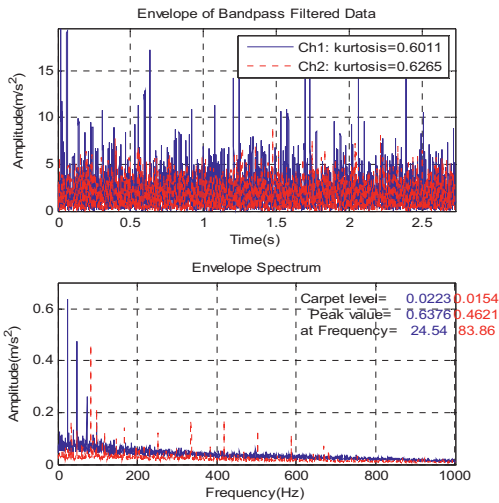


Figure 7 Envelop of bandpass filtered data and envelop spectrum

3.2 Wireless System

In this experiment, two channels of vibration signals were collected from the accelerometers mounted on bearing housing in horizontal position. The experiment was conducted under a shaft rotational speed (SRS) of 1461.7 revolutions per second (RPS), that is, 24.36 Hz, and load of 50 % with sample rate of 10000. The data is collected from two vibration sensors in time domain and the data converted from the time domain into the frequency domain using the Fourier transform. Envelope analysis is then applied which is achieved by band pass filtering and FFT calculation. These calculations are all carried out in a Matlab program. The signal acquired from the bearing has an outer race fault, shown in Figure 9 which looks complicated, with successive impulses occurring at somewhat similar spacing between one another. The RMS value of this waveform is 0.031136. Once again; the shaft rotational frequency and its multiples can be easily identified while the characteristic fault frequency of the outer race and its harmonics are obvious in the spectrum. It also shows that the frequency is around 0.876 kHz and has the highest peak with an amplitude of 0.20183 and the range of 0.05-0.2 kHz also contains large amplitude components. Figure10 shows its advantages over FFT, with clearly visible periodic components. Note that the highest peak from the bearing with the faulty outer race is 0.0048 with a frequency of 876 Hz

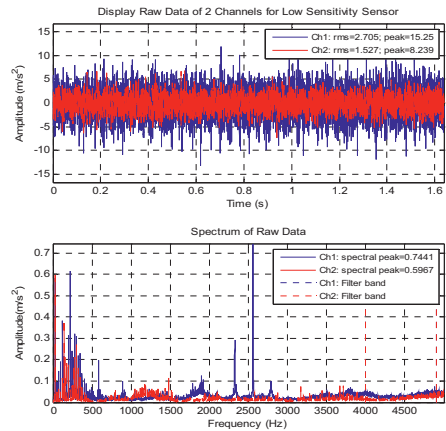


Figure 8 Wireless raw data and its spectrum

By applying envelop spectrum on the vibration signals as shown in Figures 10, it can be seen that the red spectrum has the outer race fault. In order to assess the clarity of different defects in the bearings, spectrum analysis is used as it is shown in Figures 10 for both the healthy and defective bearing. The frequency spectrum of the vibration signal from the bearing with the outer race defect peaks at 83.6 Hz, 167.2 Hz, and 249.9 Hz. The fundamental frequency for the bearing with the outer race defect is found to be 81.65 Hz.

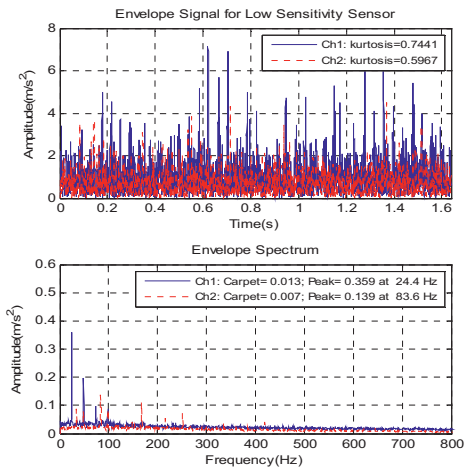


Figure9. Wireless envelop signal and envelop spectrum

4. COMPARISON OF WITH WIRE AND WIRELESS

It can be seen from figure 8 and 10 the similarity between the wired network and the wireless network is good to identify the outer race fault. In this test the sampling rate of the wired network is set at 12,000Hz so that it can match closely to the wireless network has 10,000Hz.

The difference between the wired and the wireless network is the peak value. The wired network has frequency of 83.86 where the peak value is 0.4621 which identifies the outer race fault. On the other hand the peak value from the wireless network test is 0.139 at the frequency of 83.6 because of the frequency bandwidth of wireless system is 1000Hz narrower. Therefore, the peak value acquired from the wired network is significantly greater than that of the wireless network.

5. CONCLUSION

Condition monitoring is important to industry. However, wired measurements have problems in installation. Therefore this paper investigated the wireless based condition monitoring. Wireless sensors are gaining popularity in condition monitoring applications because of their relatively low cost and ease of installation. Most wireless systems are of the IEEE 802.15.4 standard for communication system. Zigbee is a standard for Wireless Personal Area Network (WPAN) applications that require low data rates and low battery life. To evaluate this communication protocol system in condition monitoring, a

vibration node is developed by a supplier and evaluated in the lab. This evaluation is based on a bearing condition monitoring practice. Time waveform and frequency spectrum provide useful information to analyse defects in bearings. Time waveform indicates severity of vibration in defective bearings. Frequency domain spectrum identifies amplitudes corresponding to defect frequencies and enables to predict the presence of defects on the outer race. The evaluation results show that the presence of peaks in the spectrum at the second bearing, were not exact multiples of the shaft rate. This was the first indication that a bearing problem was likely to exist. The presence of peaks around these bearing tones confirmed that this vibration was in fact bearing related, as external vibration would not cause this effect and because this pattern is common to outrace faults. The presence of the same peak at 83.3Hz in the demodulated data from the bearing, therefore, confirmed that the problem was at the outrace. However, some problems with wireless network are related to the bandwidth rate.

6. REFERENCES

- [1] Bloch, H. P., (1988) Practical Machinery Management for Process Plants, Gulf Publisher.
- [2] Dilpiaz, G., and Rivola, A., (1997) Condition Monitoring and Diagnostics in Automatic Machines: Comparison of Vibration Analysis Techniques. Mechanical System and Signal Processing.
- [3] Kelly, A. (1997) Maintenance Strategy. Oxford: Butterworth-Heinemann.
- [4] Mobley, R. K., (1998) Condition Based Maintenance. In: Handbook of Condition Monitoring, Davies, A, ed.London: Chapman & Hall.
- [5] Collacott, R. A., (1997) Mechanical Fault Diagnosis and Condition Monitoring.
- [6] BS909N Wireless Gateway Available from <http://www.globalsensortech.com/bs909n-wireless-gateway.html> [Accessed 11/09/2011].
- [7] ZigBee Available from <http://en.wikipedia.org/wiki/ZigBee> [Accessed 11/10/2011].
- [8] Allan, D. Time and Frequency (Time-Domain) Characterization. Available from: <http://tf.boulder.nist.gov/general/pdf/752.pdf> [Accessed 25/09/2011].
- [9] Frequency domain methods Available from http://dip.sun.ac.za/~hanno/tw444/lesings/lesing_8.pdf [Accessed 11/09/2011].
- [10] Data acquisition Available from http://en.wikipedia.org/wiki/Data_acquisition [Accessed 18/10/2011].

Re-Design of AC Excitation Busduct based on Infrared (IR) Thermography: Condition-Based Monitoring (CBM) data analysis

Ir. S.Gopinath
 Senior Electrical Maintenance Engineer
 Jimah O&M Sdn Bhd
 2 x 700MW JEV Coal-fired Power Plant
 PT7308 & PT7309, Mukim Jimah,
 71960 Port Dickson, Malaysia
 +60165622622
 gopinath@jimahev.com

Ir. Abdul Razak b. Jamin
 Electrical Maintenance Head
 Jimah O&M Sdn Bhd
 2 x 700MW JEV Coal-fired Power Plant
 PT7308 & PT7309, Mukim Jimah,
 71960 Port Dickson, Malaysia
 +60193755790
 razak@jimahev.com

ABSTRACT

This paper presents the Infrared Thermography measurement and data analysis which lead to the Re-Design of AC Excitation Busduct. The previous AC Excitation Busduct has shown hotspots at the bolts and flanges of more than 100 degC. Further internal inspections reveal severe damage to the post insulators.

A detail study on the Re Design of the AC Excitation Busduct was conducted to overcome the hotspots problem and also to avoid major catastrophic failure to the busduct. If not attended, prolong condition could result in flashovers and lead to plant breakdown.

The outcome of the study recommends major modification work on the AC Excitation Busduct. The modification was successfully implemented at Unit20 (U20) AC Excitation Busduct. The new design Busduct has improves the temperature with less than 45 degC throughout the entire portion of the Busduct.

Keywords

AC Excitation Busduct, Infrared Thermography, Condition-based Monitoring Maintenance.

1. INTRODUCTION

Thermographic technology is commonly used for electrical equipment inspections to detect hotspot temperature. Common failures of electrical equipment is flashovers due to insulation breakdown. Main contributor for insulation breakdown is overtemperature and hotspot.

As electrical connections become loose, there is resistance for the current to flow which will cause an increase in temperature. The increase in temperature will cause hot spots resulting in insulation breakdown and components to fail.

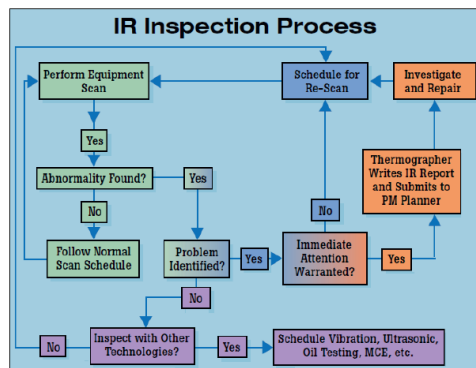
If left unchecked, heat can rise to a point that connections melt and break the circuit; as a result, fires may occur. Thermography can quickly locate hot spots, determine the severity of the problem, and help establish the time frame in which the equipment should be repair.[1]

Based on DOBLE Isolated Phase Busduct Testing with EMI Diagnostics, loose conductor connections and cracks in post

insulator will result in discharge activity and prolong condition may result in heat generation which will lead to busduct failures. [2]

Since the Thermography techniques has proof its effectiveness in monitoring heat dissipation on equipments, JIMAH has establish the IR thermography technology as part of preventive maintenance program on electrical equipment at power plant such as Circuit Breakers, Busduct, Motors and Transformers.

Baseline images on the critical equipment above has been collected using IR Thermography techniques and analyze according to the Flowchart 1 below. These baseline data will be used for reference in troubleshooting and rectification pertaining to high temperature equipment.



Flowchart 1: The flowchart demonstrates how Thermography fits into an overall maintenance program that includes other PdM technologies. (Photo courtesy Greg McIntosh, Snell Infrared Canada) [3]

On-line thermal scanning on contact terminals will identify the hot-spots and criticality of the fault (See Table 2).The hot spot is indicated by high surface temperature shown on the equipment. By attending to the hot spot will eliminate the breakdown of the system and reliability can be improved. [4]

Table 2: Criticality of Thermography survey for electrical joints [5]

SL	Criteria (Differential temperature above ambient)	Criticality Condition	Recommended Action
1	Up to 10 °C	Non-Critical	No action is needed
2	Between 10 °C to 20 °C	Less Critical	Regular Monitoring is needed
3	Between 20 °C to 40 °C	Semi Critical	Close monitoring needed. Should be attended in the next opportunity.
4	Above 40 °C	Critical	Should be attended immediately as per the severity.

2. PROBLEM STATEMENT

Jimah Power Plant generates 1400MW of electricity supply to Malaysia National Grid System. It comprises of 2x700MW of Coal Fired Steam Turbines with a sub critical Boiler located at Port Dickson, Negeri Sembilan.

A power plant has many critical equipment set up together to perform the intended function and operation. One of the critical equipment is the Non Segregated Busduct which is made up of aluminium bar conductors (busduct) for the purpose of supplying electrical power to the excitation system.

This type of busduct configuration is made of high conductivity material, aluminum alloy, supported on cast-resin post insulators. The Excitation Busduct consists of 241 pieces of post insulators. The cross sectional diagram is shown in Figure 3.

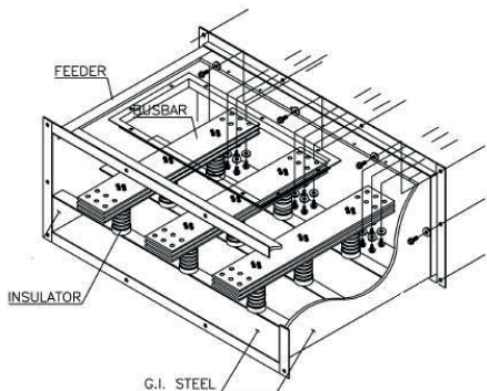


Figure 3: Cross sectional diagram of AC Excitation busduct

The AC Excitation Busduct is connected between the Excitation Transformer and the Thyristor Rectifier (AVR) cubicle. AC Excitation Busduct carries power supply with full load of 1140 VAC and 5050 AAC with 26 connecting portions/blocks from Excitation Transformer to Thyristor Rectifier board. Refer to the Diagram 4 below on AC Excitation busduct circuit configuration and Figure 5 on actual busduct location at site for better understanding.

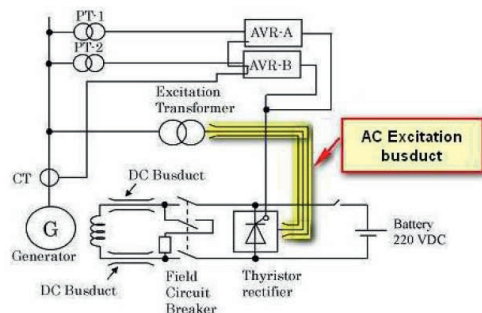


Diagram 4: AC Excitation busduct circuit configuration



Figure 5: Actual busduct location at site

Based on overall plant thermograph conducted at Unit 10 AC Excitation Busduct on 31st Oct 2008, it was reported during 100% load some of the bolts on the flanges of AC Excitation bus duct had indicated high temperature approximately around 110 deg.C. Refer to the Figure 6 below on infrared Thermography imaging.

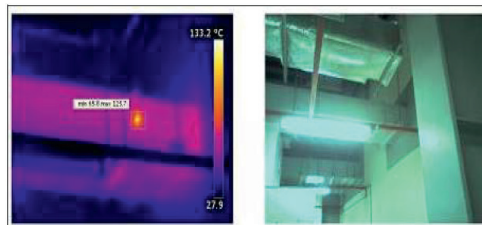


Figure 6: Defective U10 AC Excitation Busduct

In rectifying the problem, JIMAH have carried out inspection to check on the condition of AC Excitation busduct. Upon checking

on internal parts, we found that most of the post insulators were broken. (Refer to the Figure 7) on damaged post insulator inside the busduct. This similar problem had occurred on both units.



Figure 7: Damaged post insulators inside the busduct

Total 70 pieces of post insulators were found damaged. Hence, we had replaced all post insulators in the entire busduct with new post insulators (same size). However, this does not resolve the problem because we encountered the same problem during the next plant outage. The hot spot temperature issues were also not fully resolved.

3. REMEDIAL MEASURES BASED ON INFRARED THERMAL IMAGING

After further analysis and discussion, JIMAH have established few action plans to resolve the post insulator cracks and AC Excitation Busduct hot spot temperature issue.

These remedial actions took almost two (2) years to resolve the problem. The remedial measures based on infrared thermal imaging are as per below:

- Action Plan #1 – Install additional 50mmsq jumper cable on the flanges [Final result: Hot spot temperature around the jumper cables connection and post insulators damages still occurred. Thermography scan still shows high temperature hotspot]
- Action Plan #2 – Install silver plated copper plates on the flanges & change to 85mm post insulators [Final result: Hot spot temperature around the silver plated copper plates and post insulator damages still occurred]
- Action Plan #3 – Install the new Re-Design Busduct with 85mm post insulators [Final result: Successfully reduces the hot spot temperature with no post insulator damages]

3.1 Action Plan #1 - additional 50mmsq jumper cable

During the initial bolts overheating investigation on October 2008, (bus duct supplier) suspected the effect of eddy current which caused the overheating of the bolts. Hence, additional 50mmsq jumper cable and rubber were installed. However, the rectification on the overheating bolts doesn't solve the problem since post insulators still damaged and hot spot around the jumper cable connection still exist. Refer to the Figure 8 below on hot spot temperature detected around the jumper cable connection at the flanges.

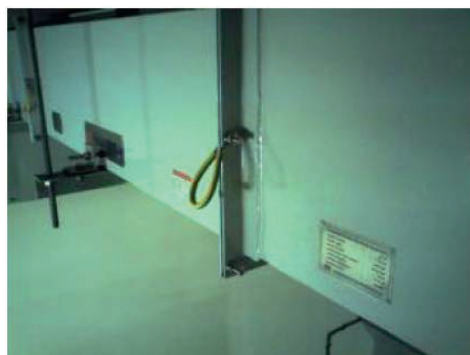


Figure 8: Hot spot temperature around the jumper cable connection

3.2 Action Plan #2 - silver plated copper plates on the flanges & 85mm post insulators

Further investigation revealed that the effect of circulation current will pass from one flange to the next adjacent flange through the connecting bolts (since there is a layer of gasket between two flanges). Besides, the root cause of the post insulator damages is due to excessive force occurred during the operation of the Bus duct. This force originated from the thermal expansion. In order to overcome this problem, we have come up with new modification on Busduct flanges connection and bigger post insulators (from 75mm to 85mm diameter size) completely. Silver plated copper plates installed on the flanges so that, circulating

current will pass through the copper plates instead of 50mmsq jumper cable on the flanges bolt connections. Refer to the figure 9 and 10 below on copper plates on the flanges connection and 85mm diameter size post insulators installed inside the busduct.



Figure 9: Copper plates on the flange connection



Figure 10: 85mm diameter size post insulators

Although provision for thermal and stress expansion has been provided to the busbar in actual application environment, the allowable expansion cannot be utilized thus the busbar expansion has been forced upon the insulator fixture. This has caused the insulator exceed the breaking strength and breakdown. Installing an 85mm diameter insulator would prevent the damage to re-occur at bus duct as it has 1.5 times greater strength than the current 75mm insulators used. [6]

Infrared thermal images revealed that the temperature of the metal plates on the flanges reduced below 60 degC. However, the modified post insulators still cracks after few months which leads to hot spot temperature rises near 100degC at flanges. Refer to the figure 11 and 12 on IR thermal gradient around silver plated copper plates on the flanges and damaged 85mm post insulator inside AC Excitation Busduct respectively.

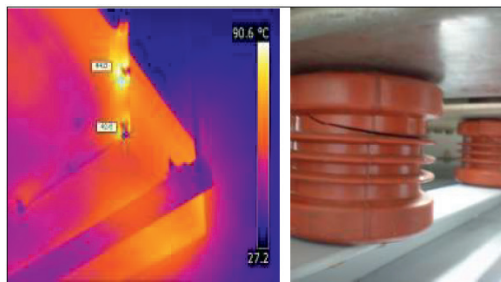


Figure 11 & 12: IR thermal gradient around silver plated copper plates on the flanges (left) and damaged 85mm post insulator inside AC Excitation Busduct (right)

3.3 Action Plan #3 – New Busduct design with 85mm post insulators

Root-cause-analysis (RCA) was conducted between HENIKWON, TOSHIBA and JIMAH Electrical Team to identify all the possible causes which could contribute to the damages of the post insulators. RCA focused on the design, raw materials, manufacturing processes and actual site condition as the possible causes. Based on the RCA and past history of insulator failures, final study was conducted on the following aspects:

- Main conductor sizing
- Temperature rise test result
- Selection of material for enclosure
- Elongation under the normal operation temperature
- Heat generating and dissipation on the conductors

Based on previous rectification and result, new AC Excitation Busduct configuration was proposed and tested. After final confirmation from all parties involved, the new Re-Design AC Excitation busduct was installed at site.

The technical parameters of the new Re-Design AC Excitation busduct are as per below:

Technical Parameters	Original Configuration	New Configuration
Rated Voltage	1.2 kV	1.2 kV
Rated current / symmetrical short circuit withstand current	5050A / 34kA (sym RMS) 2 sec	5050A / 34kA (sym RMS) 2 sec
Main Conductor size(mm)	12t * 150 (w) * 3 layers (5440mmsq)	12t * 100 (w) * 2 * 3 layers (9000mmsq)
Main Conductor material/ Finishing	Copper / tinned plated	Copper / Black coating
Enclosure size (mm)	370H * 920W	450H * 1050W
Enclosure material & Thickness size (mm)	3 sides steel & 1 side stainless (2mm)	4 sides aluminum (2mm)
Insulator size & model	Φ75 mm, JEIL Korea	Φ85 mm, JEIL Korea

Refer to the figure 13 and 14 on AC Excitation Busduct configuration of the New Re-Design and original design.



Figure 13: The New Re-Design Busduct



Figure 14: Original Design Busduct

The New Re-Design busbar configuration has two conductors per phase compared to single conductor on the existing busbar. Hence, the new configuration reduces the conductor expansion and movement because of the larger conductor size where the current is equally distributed across the 2 layer conductors which lead to lesser heat dissipation. [7]

U20 AC Excitation Busduct modification was done successfully on September 2011. Based on overall IR imaging measurement on both units, the modified U20 AC Excitation Busduct gives almost 20 degC lower temperature at enclosures and bolts on the flange compared to U10 AC Excitation Busduct. Refer to Figure 15 on IR imaging comparison between actual and modified busduct.

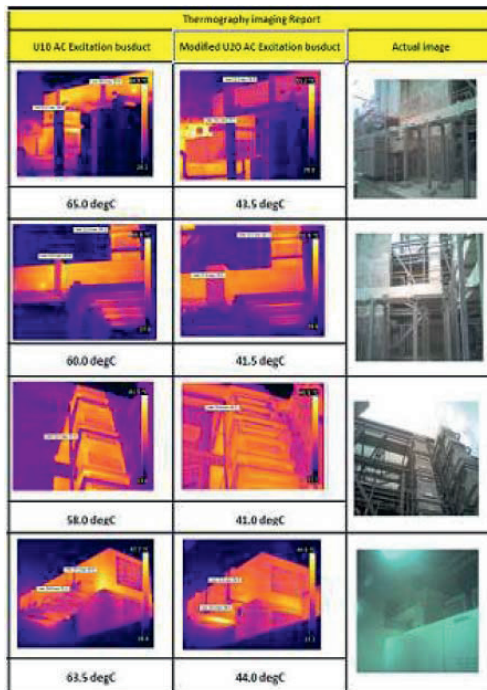


Figure 15: IR imaging comparison between actual and modified busduct.

Based on Figure 16, action plan No.1 & No.2 have not resolved the hot spot problem. After implementing action plan No.3, the AC Excitation busduct hot spot temperature had reduced from 110 degC to 45 degC with no post insulator damage. Similar modification will take place for U10 AC Excitation Busduct during plant shutdown on September 2012.

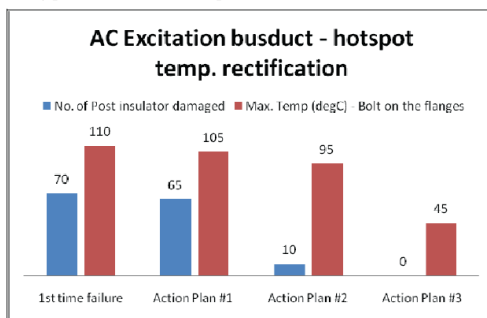


Figure 16: Comparison chart on the remedial action plan for the AC Excitation busduct hot spot temperature showing number of post insulators damage

4. CONCLUSION

As a conclusion, we find IR Thermography technique plays an important tool and method for routine electrical maintenance program in enabling us to closely monitor all electrical equipment at power plant in preventing failures or equipment breakdown due to looseness and undersized design.

5. ACKNOWLEDGMENTS

The authors wishes to acknowledge Ir. Hj. Ahmad Faiz bin Mohd Salleh (Station Manager) who has motivates us to instigate this conference paper and technical staff from JIMAH O&M, TOSHIBA and HENIKWON Cooperation for their rectification work on AC Excitation busduct at 2 x 700MW Coal-fired Jimah Power Plant at Port Dickson, Malaysia.

6. REFERENCES

- [1] V.J. Greene, "Infrared Inspection," The Locomotve, vol. 64, no. 5, pp 107-113, Spring 1985
- [2] DOBLE Tester, 'EMI testing on Insulated Phase busduct (IPB)', 2009.
- [3] J. Snell, "Improving the results of thermographic electrical inspections," ASNTFallCon~R record, Oct. 1996, John Snell & Assoc.
- [4] MIL-STD-2194, "Infrared thermal imaging survey procedure for electrical equipment," Naval Sea Systems Command, Feb. 1988
- [5] Kaplan, Herbert, "ASNT Level III Study Guide on Infrared & Thermal Testing Method"
- [6] Toshiba & Henikwon, " Incident reports of AC Excitation busduct insulators failures", 2009.
- [7] Toshiba & Henikwon, "Rectification & analysis of AC Excitation busduct – Post Insulator failures ", 2011.

Machine Health Monitoring: An Integrated Maintenance Approach

Mahantesh
Nadakatti

Division of Operation,
Maintenance and Acoustics
Luleå University of
Technology
Luleå, SE-971 87, Sweden
+46 920 49 24 83

mahantesh.nadakatti
@ltu.se

Sarbjeet Singh

Division of Operation,
Maintenance and Acoustics
Luleå University of
Technology
Luleå, SE-971 87, Sweden
+46 920 49 24 83

sarbjeet.sing
@ltu.se

Aditya Parida

Division of Operation,
Maintenance and Acoustics
Luleå University of
Technology
Luleå, SE-971 87, Sweden
+46 920 49 24 83

aditya.parida@ltu.se

Uday Kumar

Division of Operation,
Maintenance and Acoustics
Luleå University of
Technology
Luleå, SE-971 87, Sweden
+46 920 49 24 83

uday.kumar@ltu.se

ABSTRACT

Machine health monitoring in today's advanced and complex machine / plant systems have gained more prominence than before because of steep increase in machinery costs, plant investments and maintenance expenses. A breakdown in any one machine or a component in a plant could mean huge losses coupled with safety environmental threats as in case of nuclear or chemical plants. In order to efficiently manage today's maintenance functions, many methodologies are being proposed. The present paper proposes one such concept "Integrated Maintenance Model (IMM)" for efficiently integrating maintenance functions with other plant functions in order to achieve the goals and objectives of an optimum and efficient operations and maintenance. In today's highly automated plants, vast data is available both for on-line and off-line monitoring conditions and enabling appropriate decisions to be taken by a suitable decision-support mechanisms on continuous basis. The IMM model proposed in this paper discusses the ways of integrating maintenance data, maintenance decision support systems, e-maintenance, data management with suitable sub-modules along with a feedback mechanism for control. The paper outlines possible sub-modules for IMM along with their features.

Keywords

Fault-diagnosis, Performance Measurement, Integrated Maintenance Module.

1. INTRODUCTION

Maintenance is the routine and recurring process of keeping particular machine or asset in its normal operating conditions so that it can deliver its expected performance or service without causing any loss of time on account of accidental damage or breakdown. In other words, maintenance means the work that can be utilized to its full designed capacity and efficiency for maximum amount of time. The development of mechanization and automation of production systems and associated equipment, with the accompanying development of ancillary services and safety requirements, has made it mandatory for engineers to think about proper maintenance of equipment. It important to make effective use of available facilities that have been setup with high investment costs. To achieve maximum profit, it is essential to run the equipment efficiently and is only possible when the equipment or facilities are looked after properly.

Many activities of maintenance processes particularly failure detection, testing, diagnosis and repair, are knowledge intensive and experience-based tasks. Skilled maintenance engineers, apart from using test procedures and maintenance manuals, traditionally use their intuition or heuristics and understanding of how the system works to solve the problems. They must have a knowledge of how the system works and its relation with the outside world, how the system will behave when a certain subsystem fails, symptoms and failed subsystems, and finally an understanding of fault diagnosis and repairing; in essence, maintenance activity of any industry (Nadakatti et al., 2006)

The Operational Availability (OA) is expressed as: [01]

$$OA = (OT + IT) / (OT + IT + AD + RT)$$

Where:

- OA : Operational Availability
- IT : Idle Time.
- AD : Administrative or Operational Delay
- RT : Repair Time.

A lot of knowledge has been gained regarding the vibration effects, imbalance, temperature rise, noise levels, lubricant oil contaminants and their origin in a machine and so on, over the decades and have been published in various handbooks, journals and other publications. The current researchers in maintenance engineering are aiming at developing a comprehensive machine monitoring systems based on above factors, which will have a very broad data base regarding the machine health monitoring, Performance Measurement [PM], data banks, feedback systems, etc., for integrated maintenance philosophy.

All operating machineries are subjected to vibrations, wear & tear. Deterioration in the machine running condition always produces a corresponding increase in these parameters. By monitoring them carefully, it is possible to obtain information about a machine condition. As the primary source of faults in industrial machinery is most probably, a mechanical origin, it is logical to choose mechanical phenomena like vibrations, wear, etc., as the main Condition Monitoring (also known as machine health monitoring) parameter. Fault diagnosis has become one domain, where knowledge based fault diagnosis systems-research interest, has been particularly intense in the past decade. Analysis of

Condition Monitoring (also known as machine health monitoring) parameter. Fault diagnosis has become one domain, where knowledge based fault diagnosis systems-research interest, has been particularly intense in the past decade. Analysis of magnitude of machine parameters like Vibrations, Imbalance, Wear, Lubricants, etc., occurring at various frequencies or at different times during the life span of a machine can provide a great deal of information, not only about the mechanical condition of the machines in general but about the condition of specific components of the machines as well. The need of today's advanced maintenance management of a plant is that, all the maintenance activities should be integrated into organization's other activities like design, manufacturing, procurement, budgets, etc., so that, an appropriate maintenance practice, best suited for that individual plant is well in place when the plant is planned or erected. Making any modifications for existing maintenance module for integration with other activities of the plant will be very difficult, but it would be of a great asset if maintenance activities get integrated right during the design and erection stage of the plant. The aim of the present paper is to present one such conceptual model titled "Integrated Maintenance Approach"

A schematic conceptual sketch of the Integrated Maintenance model for today's highly advanced manufacturing plants are shown in Figure No.1.

2. LITERATURE REVIEW

Various techniques for maintenance engineering and management, not only help companies diagnose equipment failures but often predict them and also recommend repair strategies. The interest in using these techniques for effective use of maintenance management began very recently, leading to the development of new generation tools, techniques and suitable software specifically for this domain. Maintenance strategies like relevant condition predictor (RCP) based maintenance with improved and integrated condition monitoring can be of great help to understand, monitor and maintain the machine or a plant [2] but it is not a luxury but a necessary part of daily routine [21] in industries aspiring for success with some assumptions regarding condition based maintenance [11]. Further, a lot of research has also been done on optimization of maintenance policies, namely when to continue repairing units and when to overhaul the systems [8] which can be further integrated with design and operational systems, and consolidates some of the successful maintainability approaches to formulate an effective solution for optimized plant maintenance [9]. It has been observed that as a part of production process, maintenance programs can play a much more active role in the management of manufacturing systems. Properly designed maintenance programs can enhance the overall system performance by reducing the need for inventory, smoothing production flows and improving product quality [23]. Performance Measurement using the balanced scorecard tracks the key elements of a company's strategies from continuous improvement and partnership to teamwork and global scale [19]. The capability of automatically detecting and diagnosing the presence of localized defects in a given bearing system as well as evaluating the extent of bearing damage can be built into the a digital computer [4]. The data obtained from condition monitoring can be analysed and can provide information

for current or future maintenance decisions for predictive maintenance and operational planning [12]. Expert Systems (ES) and Decision Support Systems (DSS) are being extensively developed to eliminate the conventional programming problems. These tools can help maintenance managers in manpower planning, inventory control, prediction of PM intervals and so on. The knowledge base mainly consists of rules describing the relationship between machinery problems and the corresponding symptoms [17]. The real time expert system manages the sensor data, along with information from other plant sensors and the information can be used to advice operators for corrective actions. Several case studies have been done about the plants which have installed experts systems for detection of leaks from potentially hazardous areas at designated locations across the site, monitor alarm status and detect failures of any existing alarms connected to the terminal automation System and use the above information to detect problems early and advise site personnel [5]. Knowledge-based systems are appropriate tools for the diagnosis of faults in complex devices and that both deep and shallow knowledge have their part to play in this process. The successful implementation and evaluation of the two diagnostic knowledge-based systems are appropriate tool for the diagnosis of faults in complex hydro-mechanical devices and that they make a beneficial contribution to the business performance of the host organization [7]. An Integrated Computerized Maintenance Management (CMM) System has also been developed for integrated functioning of a manufacturing organization and to replicate the processes of a traditional ERP system, collect the real time data, explode the data against the timeline, help in balancing the use of limited resources against the timeline and finally give output in the form of preventive maintenance and corrective maintenance work orders [22]. The mobile technologies have the potential to redefine and re-engineer the conventional setting for industrial asset and maintenance management, with key characteristics being 24/7 web-based interactivity, ready access to knowledge and information, and growth of use of advanced communication networks [10].

It can be observed from the literature review that, a lot of research is being done in the field of Maintenance Engineering and Management. Integrating maintenance operations with other plant operations like design, manufacturing, budgeting, scheduling, etc., is the need of the hour. Techniques for integrating maintenance operations with all other plant activities appeared fairly recently, because of rapid developments in the fields like Information Technology and Internet. Most companies have failed to develop maintenance strategies in-line with their investments in the advanced manufacturing technologies. This has resulted in a considerable gap between the required skills, which are essential to maximize the potential benefits from these technologies and the skills, which currently exist within the maintenance sections of most industries. The application of Information Technology (IT) coupled with integration of sensors, failure predictive methodologies, Performance Measurement techniques for maintenance and operation of the industries can lead to improvement in productivity, utilization, cost control, plant safety and care for environment, health and safety of operating personnel.

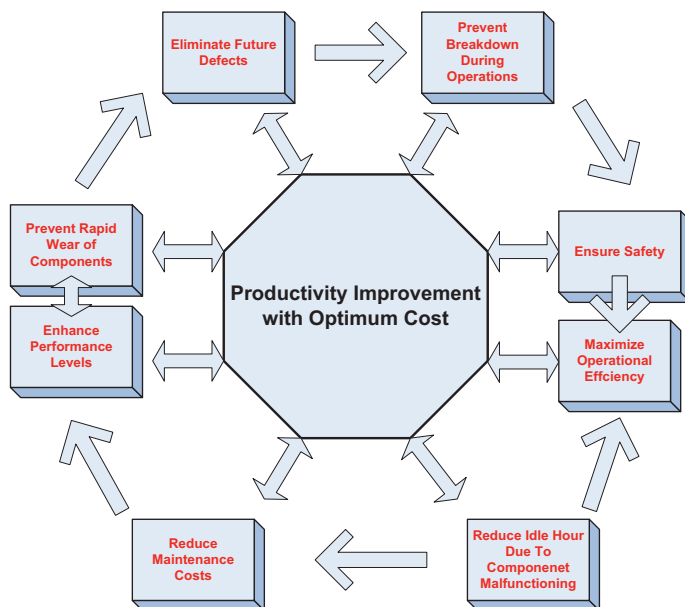


Figure No. 1. Productivity Improvement Through Optimum Asset Maintenance

3. Integrated Maintenance APPROACH

Integrated Maintenance Module (IMM) is a conceptual model for efficient maintenance and management of today's complex machine systems and plants. With rapid developments in Information Technology and high-speed Internet, the plant performance can be enhanced with inputs from sophisticated sensors embedded at suitable locations all over the plant. Software developments in areas of Artificial Intelligence (AI) like expert systems, fuzzy logic, pattern recognition, and hybrid systems make it possible to process larger volumes of data and make advanced reasoning and decision routes [13]. Integration of maintenance operations involves many key elements like: machine data, failure history, level of automation, type of machinery, type of plant, etc. The main goal of any integrated maintenance module would be to increase Availability (or reduce Down-time), Reliability, Performance and Overall Equipment Effectiveness (OEE). As can be seen from IMM Module in Figure No.1, most of the above-mentioned factors are considered for achieving these goals

3.1 IMM Model:

The IMM model consists of six sub-modules namely:

- i. Corporate / Maintenance Objectives and Goals Module
- ii. Maintenance and Performance Measurement Module
- iii. E-maintenance Module
- iv. Integrated Maintenance Module
- v. Decision Support / Decision Making Module

vi. Feedback Module

3.1.1 Module 1: Corporate / Maintenance

Objectives Module

Maintenance Objectives are derived in-line with Corporate Objectives. The main objective of any organization would be enhance productivity leading to profits with due concern about various aspects of manufacturing like safety, rules and regulations, environment, waste disposal, etc. Accordingly the main objectives of Optimum Asset Maintenance, which could result in Productivity Improvement, are shown in Figure No.1. The top management lays out the plans and budgets for current and future developments of the plant or organization.

Maintenance Objectives are then planned in accordance with organizational plans. Then suitable budgets are allocated for maintenance activities as per plans. These documentations are fed to the DATA BANK for further analysis or retrieval.

3.1.2 Module 2: Maintenance and Performance

Measurement Module:

This module consists of all those operations connected with machine / plant maintenance like different maintenance methodologies (Preventive, Predictive, Condition Based Maintenance, Lubricant Analysis, etc.). This will have support / inputs from advanced sensors for collecting data of critical parameters like machine vibrations, pressure, temperature, volume, surface roughness, mass, rate of output, lubricant level, lubricant condition, etc.

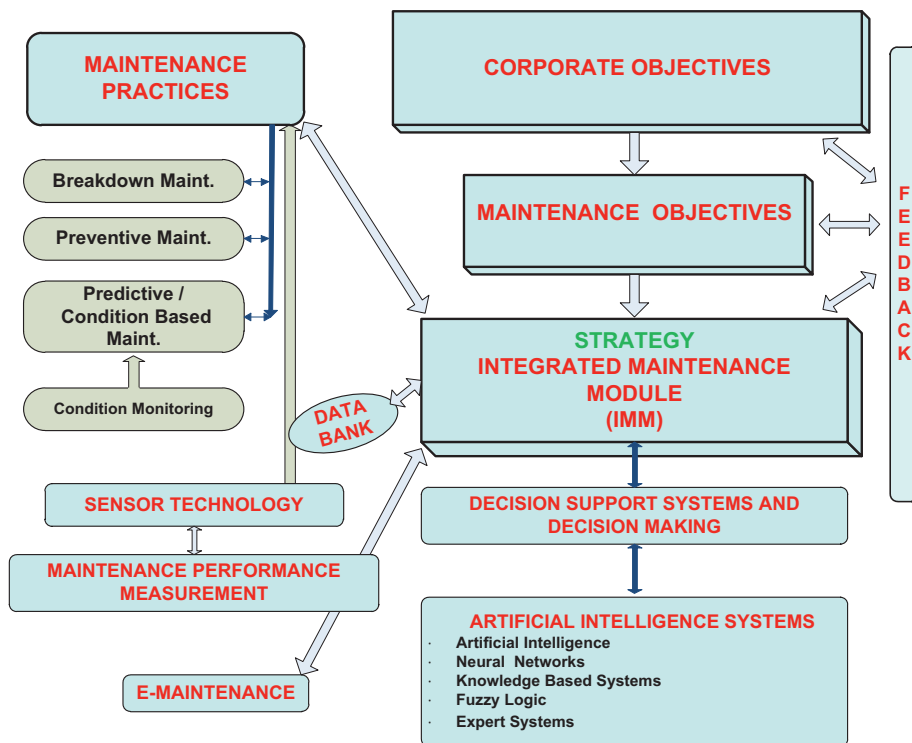


Figure 2. Integrated Maintenance Module (IMM)

Some of the on-line / off-line sensors could be FFT analysers, Vibration Pickups like Accelerometers, Pressure Gauges, Bearing analysers, Ultrasonic flaw-detectors, Thermal Sensors, etc. [12]. In order to evaluate the whole maintenance process a comprehensive Performance Measurement (PM) sub-module custom-built to the needs of organization is added to this system. The output from this module is fed to the IMM for analysing, comparing with standard data and further action. Various Decision Support Systems (DSS) are embedded in the IMM module and are used for arriving at appropriate solutions for maintenance actions.

3.1.3 Module 3: E-Maintenance Module

E-maintenance module is the support module to IMM for effective operation and maintenance of infrastructure and rolling stocks. The effectiveness in-turn depends on quality, timeliness, accuracy and completeness of information related to plant and machine degradation which can be very well stored, compiled, edited, retrieved, and transmitted on real-time basis with an e-maintenance module, for seamless integration into the industrial environment. E-maintenance module is customized for each organization as per its maintenance needs, facilities and budget. Many companies are providing ready-to-use e-maintenance software, but the flexibility offered by them may need to be

evaluated before being implemented. Even the cost of these software could be out-of-reach for many small scale plants. Another option for overcoming this situation could be develop the e-maintenance package as per one's needs. This offers many advantages like low cost, developed as per the maintenance needs of the plant with information about all the machines and structures available in that and scope for further enhancement of package without much cost. However, the decision whether to go for commercially available E-maintenance packages or develop in-house has to be carefully evaluated and finally selected.

3.1.4 Module 4: Integrated Maintenance Module (IMM)

This is the main module of the entire maintenance program of the whole organization. Each organization will have its own maintenance strategy. IMM is one such strategy, which can be used for today's highly advanced manufacturing plants. Because of complexities involved in production, maintenance, delivery schedules, and huge infrastructure, it is highly essential that vital information needs to be stored and accessed at all times because of operational severity, consequences, time-delays, stoppages, production losses, etc. IMM receives on continuous basis inputs from various other modules for comparison, data assimilation,

data processing, decision making, report generation and finally feedback and control actions [22]. This module controls the entire maintenance actions of the plant and machinery. At every level, the access to and from this module will be custom-enabled so that, persons from shop-floor to central management will get or add the information as per their expertise, with suitable data checks in place. The IMM will have flexibility for adding more modules or deleting few modules or sub-modules as per organizational needs. The IMM receives real-time data, which can be shared with users electronically. Reports are generated for further assessment, eliminates duplicate entry so that end users need not have to enter the data at various points. Coordination of large projects with multiple crews will become easier with IMM.

3.1.5 Module 5: Decision Support / Decision Making Module

This module will be the key decision making module of IMM. It will get the on-line / off-line data from the machine sensors about various parameters, evaluate their criticality, get the maintenance schedules and compare with the standard plans and accordingly make a decision that best suits the situation. In case of emergencies, it can send out alarms, warnings along with Based Systems [15], Expert Systems, etc. The decision support will be mechanism is aided by such Artificial Intelligence tools like: Fuzzy Logic, Neural Networks, Knowledge an on-line process, with data coming in from various monitoring mechanisms spread across the plant (sometimes remote locations via satellite- Remote Monitoring) as reflected in Figure 3. The decisions are taken by the state-of-the art decision support systems on real-time basis and

communicated back to appropriate modules through high-speed data transfer through intra-net, internet or via satellite for remote monitoring. Some such examples are for remote monitoring are nuclear power plants, off-shore oil platforms, aviation industry [5], FMS systems, automobile manufacturing plants, steel plants, etc. Increasing complexities of manufacturing process has made e-maintenance a necessity than a option because of its adaptability for on-line operations, data exchange and transmission, multi-point access, etc.

3.1.6 Module 6: Feedback Module

For any complex system to succeed a comprehensive feedback system is very essential. Machinery health monitoring through IMM has a feedback module, which receives and transmits the feedback about various maintenance aspects across manufacturing facilities to concerned departments. Feedback is also essential in case of evaluation of maintenance effectiveness. Maintenance experts could address any major observations made through the feedback and solutions can be arrived at. This module compares the organization's objectives with maintenance objectives and the on-line status of various maintenance activities. It sends out reports about maintenance status reports, history reports, schedules, criticality reports, breakdown reports, financial reports, performance reports, etc. The management in-turn can take suitable corrective actions leading to higher availability, reliability, and increased productivity, with direct and indirect benefits like increased profits, reduced downtime, increased safety, etc.

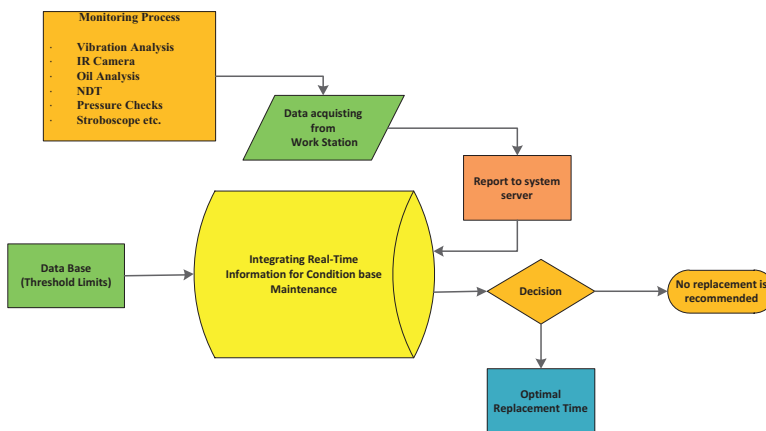


Figure 3. Decision Support / Decision Making Model

4. CONCLUSIONS

The development of mechanization and automation of production systems and associated equipment, with the accompanying development of ancillary services and safety requirements, has made it mandatory for engineers to think about optimum maintenance of equipment. It important to make effective use of available facilities that have been setup with high investment

costs. To achieve maximum profit, it is essential to run the equipment most efficiently and is only possible when the equipment or facilities are looked after properly. A high level of automation, deploying complex machines, characterizes modern industries. The objective behind Integrated Maintenance Module concept is to achieve higher reliability, availability, productivity and profits in business coupled with due concerns about safety, environment, ethics, etc., justifying the huge investments, in-

particular towards maintenance activities, needed by today's modern mechanized plants. Any downtime due to failure of some component at some location or a machine could bring the whole manufacturing process to a standstill and could result in huge losses. Machines therefore must remain in operation without any trouble. This has resulted in increased research interest in the maintenance domain. This paper proposes a conceptual model for integrating maintenance activities with overall functions of organization or a plant along with a review of current research in maintenance practices. The integration is done by top-down approach by deriving goals and objectives of plant maintenance from that of organization's and by creating individual modules one-by-one and finally integrating them. This model can be applied to major installations like nuclear plants, railways, thermal power stations, chemical plants, automobile manufacturing facilities, etc., which involve huge costs and vital impact on society in-general and organizations in-particular.

5. REFERENCES

- [1] Mishra R.C. & Pathak K. (2002), "Introduction", Maintenance Engineering & Management, PHI, pp 1-19.
- [2] Aditya P., et.al, (2008). "Relative Health Monitoring Based Maintenance for an Effective Strategic Engineering Asset Management", Proceedings of COMADEM, Prague.
- [3] Bresser, P. and Griffiths, (1993) R., "Machine Health Monitoring," SAE Technical Paper 931758, doi:10.4271/931758.
- [4] Ching Li (1987), "Online Bearing Condition Monitoring by Pattern Recognition and Matched Filtering" PhD Thesis, Dept. of Mechanical Engineering, University of Wisconsin-Madison, pp. 1-10.
- [5] Candell O, R Karim and Aditya P. (November 2011), "Development of Information System for e-Maintenance Solutions within the Aerospace Industry", International Journal of Performability Engineering, Vol.7, No. 6, pp. 583-592.
- [6] Carl Davis & Margaret Emery (June 1992). "Using Expert Systems in Plant Monitoring", C&I, pp. 38-40.
- [7] Doharty N F, Kochhar A.K. & Main R.(1994). "Knowledge Based Approaches to Fault Diagnosis : a Practical Evaluation of the Relative Merits of Deep and Shallow Knowledge", Instn. Mech Engrs, Vol. 208, ImechE, pp.39-45.
- [8] Fran Barbera, Helmut Schneider, Ed Watson (1999). "Theory and Methodology- A Condition Based Maintenance Model for a Two Unit Series System", European Journal of Operation Research, 116, pp. 281-290.
- [9] Hossam A Gabbar, Hiroyuki Yamashita, etc.(March 2003). "Computer-Aided RCM-Based Plant Maintenance Management System", Robotics & Computer Integrated Manufacturing, pp.1-10.
- [10] Jayanth Liyanage, et., al (2009). "Mobile solutions for engineering asset and maintenance management", Journal of Quality in Maintenance Engineering, Vol.15 No. 1, pp. 92-105.
- [11] Jasper Veldman, et.al.(2011). "Managing Condition-Based Maintenance Technology: A multiple case study in the process industry", Journal of Quality in Maintenance Engineering, Vol.17, No.1, pp. 40-62.
- [12] John Hill & Roger Smith (1989). "Integrating Machinery Condition Monitoring into Maintenance Management", Condition Monitoring, 2 (3), pp. 153-166.
- [13] Joze Vizintin (Dec., 2011). "Maintenance with Intelligent Diagnostic System", Proceedings of MPMM 2011 Conference, Lulea, Sweden, pp. 37-41.
- [14] Mikael Palo, et.al. (2012), "Rolling Stock Condition Monitoring Using Wheel / Rail Forces", Insight, 8/54, August 2012, pp. 451-455.
- [15] Nadakatti, M.M., et.al. (April/May 2006). IET Manufacturing Engineer, pp. 24-27.
- [16] Patick Pilat, Ilona Stouten, Jan Van Hecke, (April 2004), "How to Start Condition Monitoring", Maintworld, pp. 38-41.
- [17] Rao B.V.A. (March 1994). "Newer Techniques in Fault Logic". Fluid Power, pp. 31- 40.
- [18] Rao J.S., Jubair M., Rao C. (2003), "Condition Monitoring of Power Plants through Internet", Emerald, 14/6 2003, pp.508-517.
- [19] Rober S Kaplan and David Norton (Jan-Feb 1992). "The Balanced Scorecard-Measures That Drive Performance", Harvard Business Review, pp. 71-79.
- [20] Ruquaing Yan, Robert X Gao, (2007), "Approximate Entropy as Diagnostic Tool for Machine Health Monitoring", Mechanical Systems and Signal Processing, 21, pp. 824-839.
- [21] Siladitya Ghosh (Jul-Sep. 1995). "Maintenance Management- A Quest", Productivity, Vol.36, No.2, pp. 338-340.
- [22] Shankar Chakarborthy & Dr. Diganta Tah (Dec 2004). "Development of An Integrated Computerized Maintenance Management (CMM) Systems", Manufacturing Technology Today, pp. 3-7.
- [23] Xiao-Gao Liu, Villiam Makis, etc.(Jul-Sep. 1993). "Role of Maintenance in Manufacturing Management", Productivity, Vol.34, No.2, pp. 257-266.

Numerical Simulations of Effects of Faults in a Vertical Axis Wind Turbine's Performance

Atif Shahzad
 School of Computing and Engineering
 University of Huddersfield
 Huddersfield, UK
 HD1 3DH
 (+44) 1484 471131
 u1178899@hud.ac.uk

Taimoor Asim
 School of Computing and Engineering
 University of Huddersfield
 Huddersfield, UK
 HD1 3DH
 (+44) 1484 471138
 taimoor.asim@hud.ac.uk

Kyoo-Seon Park
 School of Computing and Engineering
 University of Huddersfield
 Huddersfield, UK
 HD1 3DH
 (+44) 1484 473531
 kyoo-seon.park@hud.ac.uk

Suman Pradhan
 School of Computing and Engineering
 University of Huddersfield
 Huddersfield, UK
 HD1 3DH
 (+44) 1484 473651
 s.r.pradhan@hud.ac.uk

Rakesh Mishra
 School of Computing and Engineering
 University of Huddersfield
 Huddersfield, UK
 HD1 3DH
 (+44) 1484 473263
 r.mishra@hud.ac.uk

ABSTRACT

Renewable sources of energy are being developed globally to overcome the present excessive dependence on fossil fuels. Wind energy is one of the important sources of renewable energy. Considerable amount of research is being carried out on the innovative designs for optimal performance of wind turbines. Furthermore a lot of research is being carried out on maintenance and condition monitoring of such systems. Torque output is one of the most important parameters in analysing the performance of a turbine; which in turn depends on a number of factors including the structural health and the performance of each blade. Cracks in a wind turbine blade affect the aerodynamic profile of the blade and consequently flow field around it, and may cause vibration in the blade further affecting its performance. In this paper Computational Fluid Dynamics (CFD) based technique has been used to study the effect of the presence of cracks in the blades on the torque output of Vertical axis wind turbine (VAWT). For this purpose, different cracks configurations have been simulated and results analysed which indicate variations in the amplitude of the torque output of the turbine due to the presence of cracks.

Keywords

Computational Fluid Dynamics, Vertical Axis Wind Turbine, Tip Speed Ratio, Torque, Power.

Nomenclature

r	Radius of VAWT (m)
w	Angular velocity (rads/sec)
v	Linear velocity (m/sec)
P	Power Output from VAWT (W)
T	Torque Output from VAWT (N-m)

1. INTRODUCTION

Fossil fuels' depletion, rising cost of fuel prices, CO₂ emissions and nuclear disasters, such as the recent one in Japan, have made renewable energy sources increasingly important. Wind energy is being considered a potential candidate in present climate. The global investment in renewable energy is increasing exponentially. Total current installed wind capacity is nearly 238GW. In 2011 more than 40GW of wind energy systems were installed, with China and India leading the contribution by sharing 50% of the installed capacity. Europe installed wind energy systems worth 10GW in 2011 while UK installed nearly 1.3GW which is 3.2% of global installed capacity [4]. In order to generate 15% of its energy from renewables by 2020, as required by the EU directive, the UK has set the target for wind energy to contribute in the range of 28 to 31GW of installed capacity [3].

Realising the potential and benefits of the wind energy, considerable amount of research is being carried out on the innovative designs for optimal performance of wind turbines with the focus on both centralized and decentralized harnessing of wind energy. Two most common designs being used are Horizontal axis wind turbines (HAWT) and Vertical axis wind turbines (VAWT). HAWT are more efficient as compared to VAWT but require good quality wind energy. In urban areas where wind is inconsistent and highly fluctuating, VAWT is more beneficial due to its low starting torque characteristics as well as other advantages like in-expensive to build and simple design [5][7].

The important performance parameters of VAWTs, as mentioned by Gareth et. al. [1], are the tip speed ratio (TSR) and the torque output. TSR is the ratio between the rotational speed of the tip of the blade and the actual velocity of the wind.

$$\lambda = \frac{r * w}{v} \dots\dots\dots (1)$$

where r is the radius of the VAWT, w is the angular velocity and v is the linear velocity. Torque output of the wind turbine has a significant impact on the total power output of the turbine.

$$P = w * T \quad \dots\dots\dots (2)$$

where P is the power and T is the torque output.

The overall output of the wind turbines depends on several geometric, flow and fluid parameters. The performance of each blade contributes towards the overall torque output of the turbine. In case the shape of the blades gets distorted, significant variations in the performance output of the VAWT could be expected. Cracks in wind turbine blades affect the aerodynamic profile of the blade and consequently flow field around it. In adverse conditions these cracks may cause vibration of the blade also, further affecting its performance. This study focuses on the investigation of performance characteristics of a wind turbine when one of its blades starts to get affected by initiation of a crack. Various operating conditions have been numerically simulated using computational fluid dynamics.

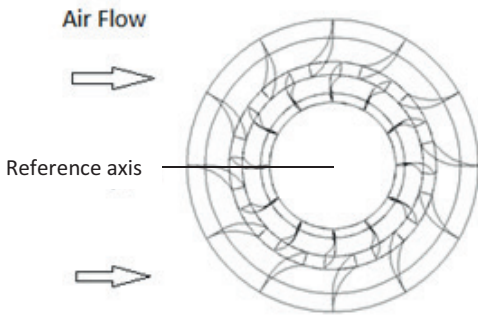


Figure 1. 3D model of the VAWT.

2. NUMERICAL MODELLING

The performance output of an in-house fabricated VAWT (shown in figure 1) has been numerically analysed for various faulty blade configurations. This VAWT has a diameter of 2m and a height of 1m. The geometry of the VAWT features 12 equally spaced rotor and stator blades respectively. Cracks of various sizes in one of the rotor blades have been generated in order to analyse the effect of the presence of faults in the VAWT on the overall performance output of the VAWT. The details of the faults have been summarised in table 1 while figure 2 shows the geometric models of the VAWTs with the faults.

Table 1. VAWT Configurations

Condition	Fault/Defect in the VAWT
Condition 1	Healthy (Non – defective)
Condition 2	25mm crack
Condition 3	50mm crack
Condition 4	100mm crack

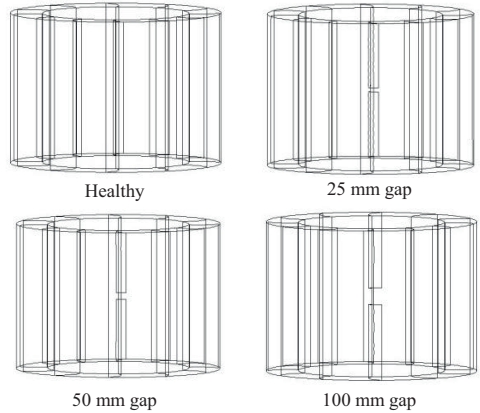


Figure 2. Various Rotor blades' configurations.

Commercial CFD package Ansys 13.0 has been used to numerically simulate the flow in the vicinity of the VAWT. The geometric details of the flow domain, encompassing the VAWT, have been shown in figure 3. 4 m/sec of air flow velocity has been specified at the inlet boundary of the domain whereas the outlet of the flow domain is assumed to be at the atmospheric pressure. The other sides of the flow domain have been specified as stationary walls with no-slip boundary conditions. k-ε turbulence model has been shown to resolve the steady-state turbulent parameters in the flow domain with reasonable accuracy [2] and hence has been chosen for analysis in the present study. Sliding mesh technique as mentioned by Park et. al [6] has been used to rotate the blades with respect to the central axis of the turbine at an angular velocity of 1.143 rad/sec such that the TSR of the VAWT is 0.2.

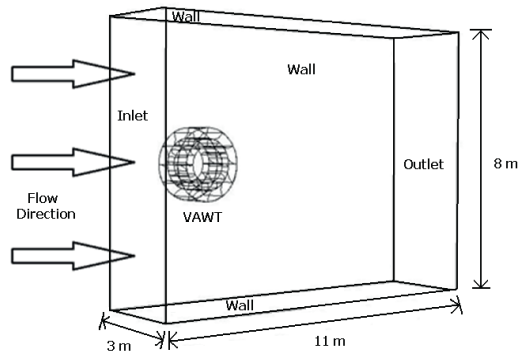


Figure 3. Flow domain encompassing the VAWT.

Three dimensional Navier Stokes equations have been numerically solved in an iterative manner to predict the flow structure in the vicinity of the VAWT for every 3^o rotation of the rotor blades. During the initial revolutions of the VAWT, significant changes in the flow structure have been observed due to the numerical diffusion. The flow structure within the flow domain has been constantly monitored. The non-uniformities in the predicted flow fields die out from 4th revolution onwards (see table 2) and hence the solution becomes statistically steady.

3. RESULTS AND ANALYSIS

Numerically converged solutions have been used to analyse the effect of various parameters on the output characteristics of the vertical axis wind turbine. To ensure that the results obtained are independent of the mesh being used for the analysis purposes, a mesh independence study has been conducted.

3.1 Mesh Independence

In order to capture the small scale features of the flow, the flow domain needs to be subdivided into small parts, known as mesh elements. To obtain fairly accurate results, the flow variables needs to be independent of the size of these mesh elements. In the present study, the flow domain has been subdivided into one and two million mesh elements respectively. The results for the average torque output, for each revolution of the VAWT having 25mm crack, have been summarised in table 2.

Table 2. Convergence and mesh independence for 25 mm crack condition

Revolution of VAWT	1x10 ⁶ Mesh Elements		2x10 ⁶ Mesh Elements	
	Average Torque (N-m)	Diff. (%)	Average Torque (N-m)	Diff. (%)
1 st	10.545		11.051	
2 nd	10.192	3.348	10.684	3.321
3 rd	10.192	0.000	10.686	0.019
4 th	10.193	0.010	10.687	0.009
5 th	10.194	0.010	10.687	0.000

On average, the difference in the average torque output from the VAWT, for both the meshes being used, is 4.8%. The mesh with two million elements yields fairly accurate results and hence has been chosen for further analysis.

3.2 Healthy state VAWT

The results presented hereafter correspond to the 5th revolution of the VAWT as it has reached a statistically steady state. Figure 4 depicts the instantaneous torque output of the healthy state VAWT. The instantaneous torque has been normalised with the average torque for that revolution of the VAWT.

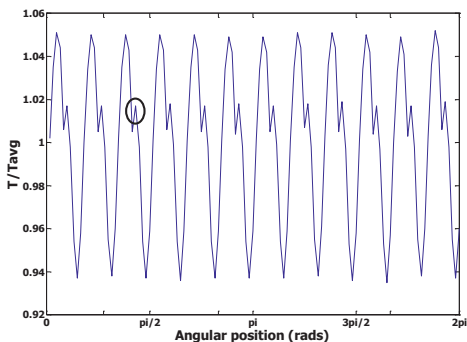


Figure 4. Instantaneous Torque output from Healthy VAWT.

It can be seen that the normalised average torque output during one revolution of the VAWT is cyclic. The average distribution of

torque is same for all the blades. Highest peaks refer to the maximum torque output when the rotor blades are in line with the stator blades, making uniform passages for the flow of air. The maximum normalised average torque output is 1.051. The lower peaks (black circle in figure 4) correspond to that orientation of the VAWT when the rotor blades are in between the two stator blades, making two passages for the flow of air. The minimum values of the normalised average torque is 0.93 which corresponds to that orientation of the VAWT when the rotor and stator blades make non-uniform passages for the flow of air, hence blocking the flow and offering greater resistance.

3.3 Faulty VAWTs

In order to analyse the VAWT under various faulty conditions, three different crack configurations have been generated in one of the rotor blades. Detailed analyses of the performance output of the VAWT under these faulty conditions have been presented below.

3.3.1 25mm crack

Figure 5 depicts the variations in the normalised torque output for one complete revolution of the VAWT having a 25mm crack in one of its rotor blades. Although the torque variation is cyclic, the torque output values are considerably different from that observed for the healthy state.

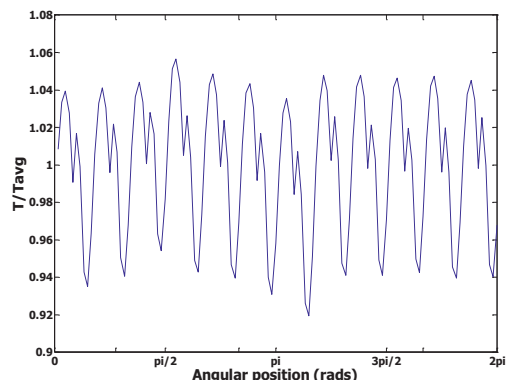


Figure 5. Instantaneous Torque output from VAWT with 25mm crack.

An increase of 0.44% in the maximum normalised torque output, compared to healthy state VAWT, has been noticed. The maximum normalised torque output is 1.056 when the cracked blade is positioned at 99^o degrees from the reference position shown in figure 1. A decrease of 1.6% in the minimum normalised torque output, compared to healthy state VAWT, has been noticed. The minimum normalised torque output is 0.919 when the cracked blade is positioned at 207^o from the reference position. The variations of the amplitude of the normalised torque output from the VAWT having a 25mm crack in one of its rotor blades is 17% higher compared to the healthy state VAWT. This variation in the amplitude of the normalised torque output leads to structural instabilities, such as vibrations, in the VAWT; hence degrading the performance output from the VAWT and adversely affecting its structural integrity and remaining useful life.

3.3.2 50mm crack

To investigate the effect of progressive increase in crack size within a blade, simulations have been carried out on a VAWT having a 50mm crack in one of its rotor blades. It is evident from

figure 6 that the torque variations for 50mm wide crack are similar to that observed in 25mm crack case. However, the normalised torque values are significantly different. It can be clearly seen that the variations in the normalised torque output are higher as compared to 25mm crack.

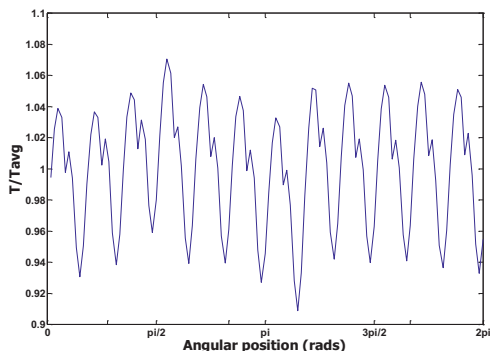


Figure 6. Instantaneous Torque output from VAWT with 50mm crack.

An increase of 1.79% in the maximum normalised torque output, compared to healthy state VAWT, has been noticed. The maximum normalised torque output of 1.07 occurs when the cracked blade is positioned at 99° degrees from the reference position shown in figure 1. A decrease of 2.77% in the minimum normalised torque output, compared to healthy state VAWT, has been noticed. The minimum normalised torque output of 0.908 occurs when the cracked blade is positioned at 207° from the reference position. The variations or the amplitude of the normalised torque output from the VAWT having a 50mm crack in one of its rotor blades is 38% higher compared to the healthy state VAWT.

3.3.3 100mm crack

To further quantify the effect of crack size, numerical simulation has been carried out with one of the blades of the VAWT having a crack of size 100mm. Figure 7 depicts the variations in the normalised torque output for one complete revolution of the VAWT having a 100mm crack.

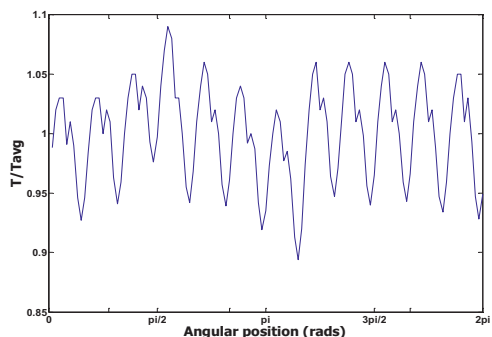


Figure 7. Instantaneous Torque output from VAWT with 100mm crack.

An increase of 3.31% in the maximum normalised torque output, compared to healthy state VAWT, has been noticed. The maximum normalised torque output of 1.086 is obtained when the

cracked blade is positioned at 99° degrees from the reference position shown in figure 1. A decrease of 4.38% in the minimum normalised torque output, compared to healthy state VAWT, has been noticed. The minimum normalised torque output of 0.893 is obtained when the cracked blade is positioned at 207° from the reference position. The variations in the amplitude of the normalised torque output from the VAWT having a 50mm crack in one of its rotor blades is 64.8% higher compared to the healthy state VAWT.

The above discussion clearly indicates that the presence of a crack in the blade significantly affects the instantaneous torque output of the VAWT. The phenomenon is further depicted in figure 8 which shows the normalised torque output for each configuration of the VAWT during one complete revolution.

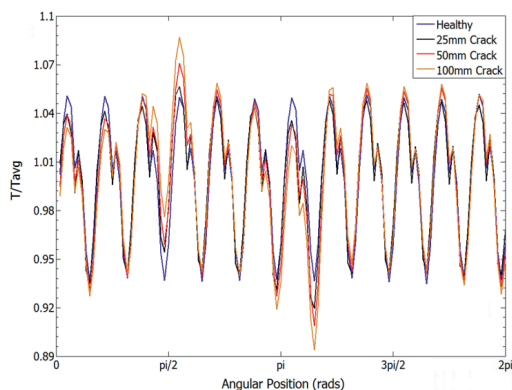


Figure 8. Instantaneous Normalised Torque outputs from various configurations of the VAWT.

Table 3 summarises the results presented here. Since average torque output obtained from the numerical for one complete revolution corresponding to each condition doesn't differ significantly, more information about the local flow field is required.

Table 3. Torque outputs from various configurations of the VAWT

Condition	Minimum Torque (Tmin) (N-m)	Maximum Torque (Tmin) (N-m)	Average Torque (Tavg) (N-m)
Healthy	0.934	1.051	1.000
25mm crack	0.919	1.056	1.000
50mm crack	0.908	1.070	1.000
100mm crack	0.893	1.086	1.000

To analyse the effects on flow field variables due to the presence of a crack in the blade, the velocity vectors have been shown in figure 9. It clearly shows that local velocities are much higher for healthy blade VAWT configuration as compared to the faulty blade VAWT configuration. The trend is the same for all the crack sizes.

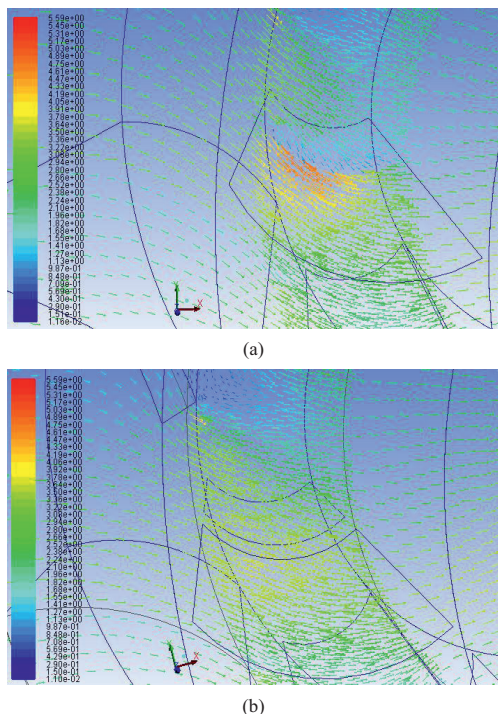


Figure 9. Velocity vectors: (a) Healthy VAWT (b) VAWT with 100mm crack.

Table 4 shows a clear correlation between the gap size and the percentage increase in the difference between maximum and minimum torque values. It can be clearly seen that as the size of the crack in the blade increases, the difference between maximum and minimum torque values increases.

Table 4. Amplitude of Normalised Torque outputs

Condition	Tmax – Tmin (N-m)	Diff. w.r.t. Healthy VAWT (%)
Healthy	0.117	
25mm crack	0.137	17.09
50mm crack	0.161	38.25
100mm crack	0.192	64.85

The presence of a 25mm crack in one of the rotor blades increases the difference between maximum and minimum torque values by 17.09%. A further increase in the crack size to 50mm further

increases the difference between maximum and minimum torque values by 17.51%. A crack size of 100mm further increases the difference between maximum and minimum torque values by 19.25%. On average, an increase of 18.38% has been observed as crack becomes double in size.

The above result highlights that increase in the value of Tmax – Tmin is an indication of increasing crack size in one of the blades of the turbine. This parameter can further be tuned to clearly isolate the crack related faults in a vertical axis wind turbine both qualitatively and quantitatively. This parameter can be embedded on any model based diagnostic system for continuous monitoring of wind turbine systems.

4. CONCLUSIONS

The paper clearly highlights usefulness of computational fluid dynamics in simulating fault related effects in wind turbines. For this purpose various wind turbine configurations have been studied and analysed to investigate the behaviour of the torque output for various crack sizes. It has been seen that during one complete revolution of the turbine, maximum torque is generated at a specific position of the rotor blade with respect to the axis of incoming air flow. The results indicate that there is a direct correlation between the gap size and the amplitude of the torque at specific orientations of the cracked blade. Furthermore crack size has a very small effect on the average overall torque output of the turbine. The difference between maximum and minimum torque values increases during one complete revolution of the turbine when the gap size increases. Hence, Tmax – Tmin can be used as a tool to diagnose the presence and size of cracks in the blade.

5. REFERENCES

- [1] Colley, G. and Mishra, R. (2011). “Computational flow field analysis of a Vertical Axis Wind Turbine”, In proceedings of the International Conference on Renewable Energies and Power Quality, Las Palmas, Gran Canaria.
- [2] Colley, G. and Mishra, R. (2010). “Effect of rotor blade position on Vertical Axis Wind Turbine performance”, In proceedings of the International Conference on Renewable Energies and Power Quality (ICREPQ’10) Granada, Spain.
- [3] DECC (2011). *UK Renewable energy Roadmap*. Available at: <http://www.decc.gov.uk/>, accessed: 20 September 2012.
- [4] GWEC (2011). *Global Wind Report*. Available at: <http://www.gwec.net/>, accessed: 28 September 2012.
- [5] Manwell, J.F. Megowan, J.G. and Rogers, A.L. (2009). *Wind energy explained; theory, design and application*. Second edition, John Wiley & Sons Ltd., Chichester, UK.
- [6] Park, K.S. Asim, T. and Mishra, R. (2012). “Computational Fluid Dynamics based fault simulations of a Vertical Axis Wind Turbine”. *Journal of Physics: Conference Series*, 364.
- [7] Rohatgi, J. and Barbezier, G. (1999). “Wind turbulence and atmospheric stability – their effect on wind turbine output”. *Journal of Renewable Energy*, 16 (1-4), pp. 908-911.

Computational Fluid Dynamics based Performance Optimisation of Vertical Axis Marine Current Turbines

Kyoo-Seon Park
School of Computing and Engineering
University of Huddersfield
Huddersfield, UK
HD1 3DH
(+44) 1484 473531
kyoo-seon.park@hud.ac.uk

Taimoor Asim
School of Computing and Engineering
University of Huddersfield
Huddersfield, UK
HD1 3DH
(+44) 1484 471138
taimoor.asim@hud.ac.uk

Rakesh Mishra
School of Computing and Engineering
University of Huddersfield
Huddersfield, UK
HD1 3DH
(+44) 1484 473263
r.mishra@hud.ac.uk

Atif Shahzad
School of Computing and Engineering
University of Huddersfield
Huddersfield, UK
HD1 3DH
(+44) 1484 471131
u1178899@hud.ac.uk

Gilbert Mirangwanda
School of Computing and Engineering
University of Huddersfield
Huddersfield, UK
HD1 3DH
U0865788@hud.ac.uk

ABSTRACT

Rapid decrease in the fossil fuels in the last couple of decades has stirred the researchers to find alternative sources for the production of power. Wind and tidal energies are the two most promising alternatives to the fossil fuels. While most of the recent research has been conducted on developing electro-mechanical systems for power production from wind energy, the research regarding the use of the tidal energy for power production is severely limited. In the present study, performance characteristics of an in-house built Darrieus type Vertical Axis Marine Current Turbine have been numerically simulated. An effort has been made to analyse and understand of the complex flow phenomenon occurring in the vicinity of such turbine. Furthermore, the optimisation study has been included for various flow configurations. It has been shown that the optimum operating condition of the vertical axis marine current turbine occurs at a tip speed ratio of 0.17 when the power production from the turbine is at its maximum.

Keywords

Vertical Axis Marine Current Turbine, Tip Speed Ratio, Torque Output, Power Output, Computational Fluid Dynamics.

Nomenclature

r	Radius of VAMCT (m)
v	Linear velocity (m/sec)
ω	Angular velocity (rads/sec)
T	Torque Output from VAMCT (N-m)
P	Power Output from VAMCT (W)

1. INTRODUCTION

Marine Current Turbines (MCTs) convert the kinetic energy of the tidal waves into useful energy forms. Two common types of MCTs are Horizontal Axis and Vertical Axis Marine Current Turbines. The principle of operation of both these types is the same as for Horizontal Axis and Vertical Axis Wind Turbines i.e. HAWT and VAWT. The use of Darrieus type Vertical Axis Marine Current Turbines (VAMCTs) provides several advantages over Horizontal Axis Marine Current Turbines (HAMCTs) such as the low starting torque, quite operation and insensitivity to the angle of incident flow. Hence, VAMCTs are better suited to extract power from tidal energy.

Two MCTs of 1 MW capacity were installed in June 2008 by SeaGen in Strangford Narrows, Northern Islands. It has been reported by Dai et. al. [3] that the dominance of the HAMCT over VAMCT is not as pronounced in the hydrokinetic energy generation as it is for wind energy. Researchers have been trying to optimise the performance of VAMCTs by modifying the geometric features of such turbines. Most of the on-going research is benefiting from the commercially available Computational Fluid Dynamics packages in order to analyse the complex fluid flow phenomenon in the vicinity of VAMCTs.

Yang et. al. [7] have numerically simulated a two dimensional Hunter Turbine using CFD and have conducted optimisation studies. They have further carried out laboratory based experiments using Particle Image Velocimetry for various rotor blade configurations [8]. Li et. al [4] have numerically simulated a VAMCT having four rotor blades. Through this study they could understand complex flow phenomena with in rotors. Bai et. al. [1] has used CFD to simulate the effects of an array of VAMCTs in an ocean bed and found effect of arrays on power capture. Paillard et. al. [5] has carried out some preliminary CFD based investigations on a VAMCT. Turnock et. al. [6] has numerically simulated an array of MCTs to analyse the wake of such turbines.

In the present study a novel in-house built Darrieus type VAMCT has been numerically simulated using a commercial CFD package. The VAMCT consists of 12 rotor and stator blades respectively where the stator blades are flat plates whereas the rotor blades have been curved at 28.2°. Figure 1 shows the geometry of the VAMCT that has been used in the present study.

The important performance parameters of vertical axis turbines, as mentioned by Colley et. al. [2], are the tip speed ratio (TSR) and the torque output. TSR is the ratio between the rotational speed of the tip of the blade and the actual velocity of the wind.

$$\lambda = \frac{r * \omega}{v} \dots\dots\dots (1)$$

where r is the radius of the vertical axis turbine, ω is the angular velocity and v is the linear velocity. Torque output of the wind turbine has a significant impact on the total power output of the turbine.

$$P = \omega * T \dots\dots\dots (2)$$

where P is the power and T is the torque output.

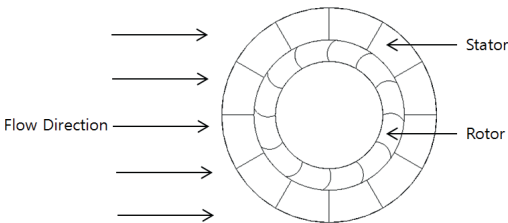


Figure 1. Vertical Axis Martine Current Turbine Model.

2. NUMERICAL MODELLING

The performance output of the VAMCT has been numerically analysed for various flow configurations. These flow configurations correspond to the tip speed ratios of 0.01, 0.05, 0.1, 0.2 and 0.4. A Commercial CFD package has been used to numerically simulate the flow in the vicinity of the VAMCT. The VAMCT has a diameter of 2m and a height of 1m. The geometric details of the flow domain, encompassing the VAMCT, have been shown in figure 2. 1 m/sec of water flow velocity has been specified at the inlet boundary of the domain whereas the outlet of the flow domain is assumed to be at zero gauge pressure. The operating pressure within the flow domain has been specified such that the VAMCT is assumed to be operating at a depth of 1 m below the surface of water. Hence, the operating pressure that has been specified to the solver is 111117 Pa. The sides of the flow domain have been specified as walls with no-slip boundary conditions and moving in the direction of the flow at 1 m/sec such that the generation of the boundary layer at these walls can be neglected. Two equation k-ε turbulence model has been shown to resolve the turbulent parameters in the flow domain with reasonable accuracy [9] and hence has been chosen for analysis in the present study. Sliding mesh technique as mentioned by Park et. al [10] has been used to rotate the blades with respect to the central axis of the turbine.

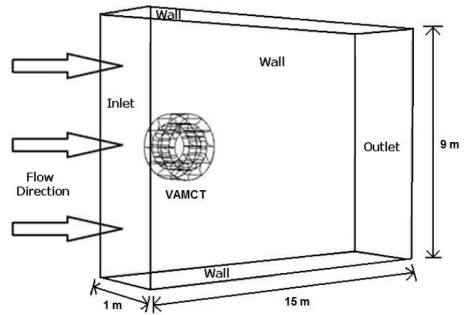


Figure 2. Flow domain encompassing the VAMCT.

Three dimensional unsteady Navier Stokes equations have been numerically solved in an iterative manner to predict the flow structure in the vicinity of the VAMCT for every 3° rotation of the rotor blades. During the initial revolutions of the VAWT, significant changes in the flow structure have been observed due to the numerical diffusion. The flow structure within the flow domain has been constantly monitored. The non-uniformities in the predicted flow fields die out from 2nd revolution onwards and hence the solution becomes statistically steady.

3. RESULTS AND ANALYSIS

Numerically converged solutions have been used to analyse the effect of various parameters on the output characteristics of the vertical axis marine current turbine. To ensure that the results obtained are independent of the mesh being used for the analysis purposes, a mesh independence study has been conducted. The results suggests that a mesh of two million elements is capable enough to capture the small scale flow related features with reasonable accuracy and hence has been chosen for analysis in the present study.

3.1 Flow Field Analysis

Figure 3 shows the velocity field in the vicinity of the VAMCT at TSR = 0.01. It can be seen that near the stator blades the velocities are very small because of the no-slip boundary condition. However within the stator blade passages the velocities increase. Within the rotor passages the velocity field is found to be non-uniform. On the rear of the VAMCT, velocities are much smaller as compared to the front end. Similar trends have been found at other TSR values.

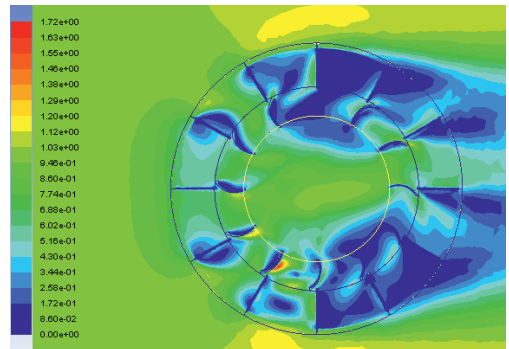


Figure 3. Variations in the Velocity magnitude in the vicinity of the VAMCT at TSR = 0.01.

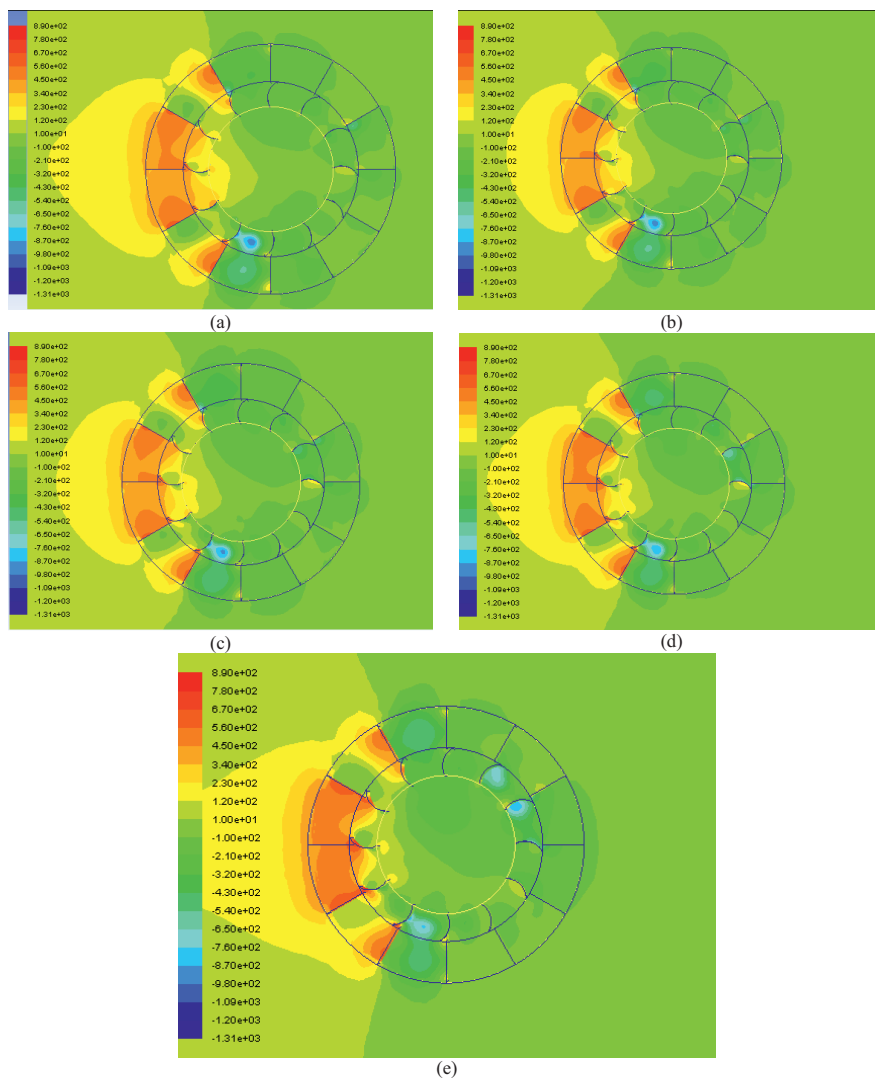


Figure 4. Variations in the Static Pressure in the vicinity of the VAMCT (a) TSR = 0.01 (b) TSR = 0.05 (c) TSR = 0.1 (d) TSR = 0.2 (e) TSR = 0.4.

In order to further analyse the flow field in the vicinity of the VAMCT, static pressure distributions for various TSRs have been shown in figure 4. The static pressure contours have been plotted against the same scale for accurate comparison at various TSRs. It can be seen that the areas of high velocities corresponds to low pressures and vice versa. Hence, high pressure region exists at the front end of the VAMCT and a comparatively low pressure region exists at rear end of the VAMCT. Furthermore, the static pressure distribution within the rotor passages and within the core region of the VAMCT is highly non-uniform.

3.2 Performance Characteristics

In order to analyse the performance of the VAMCT, instantaneous torque outputs for various TSRs under consideration have been plotted in figure 5 for one complete revolution of the VAMCT. The results presented here correspond to the third revolution of the VAMCT when the solution has become statistically steady and all the variations in the performance outputs have died out. It can be seen that the torque output from the VAMCT shows a trend of alternative peaks and valleys corresponding to different rotor blades' angular position. The peaks in the outputs are formed when the rotor blades are in-line with the stator blades hence forming large uniform passages for the flow of water. Whereas, the valleys occur when the rotor blades are positioned such that

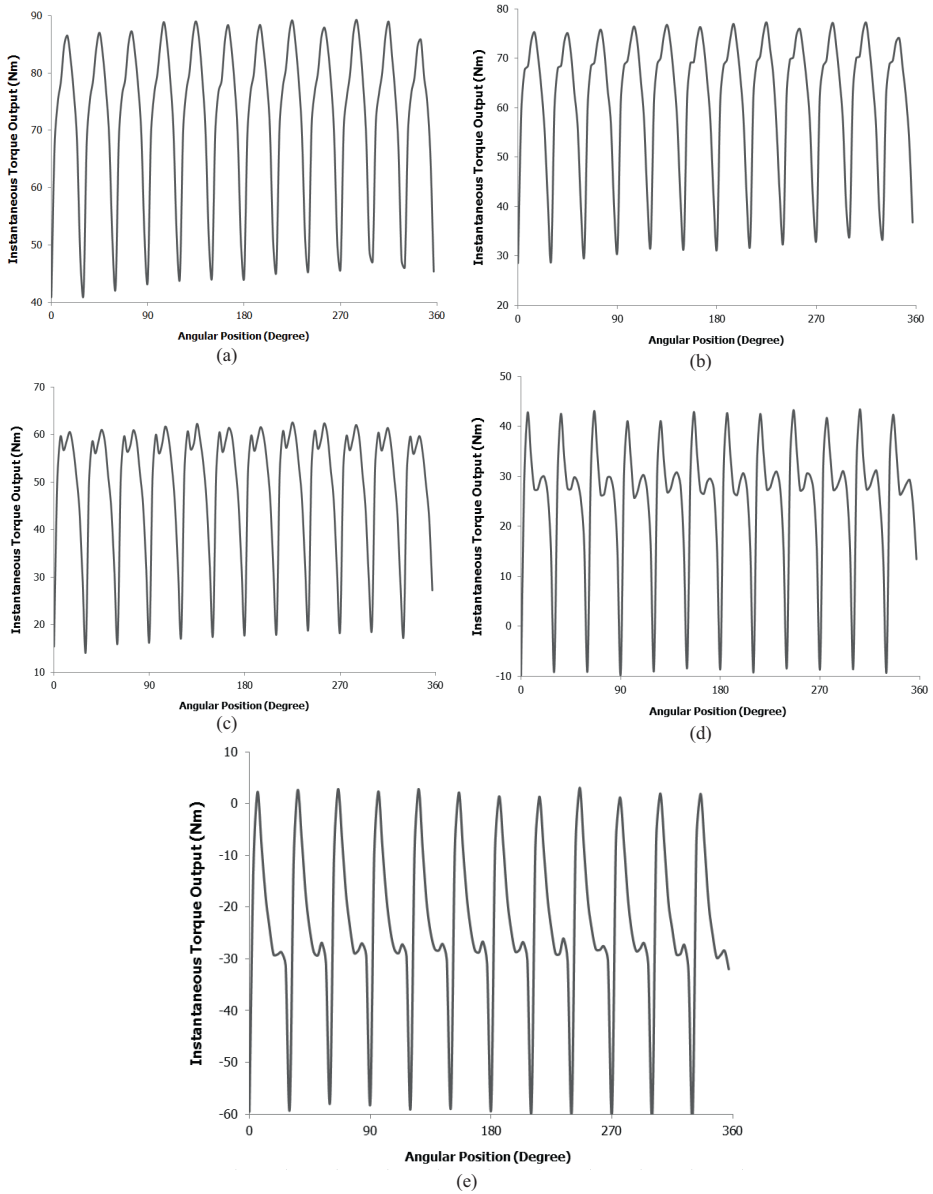


Figure 5. Variations in Instantaneous Torque Output in one revolution of the VAMCT (a) TSR = 0.01 (b) TSR = 0.05 (c) TSR = 0.1 (d) TSR = 0.2 (e) TSR = 0.4.

there exists non-uniform passages in the rotor section of the VAMCT, leading to non-uniform pressure distribution around the rotor blades, This leads to vibrations in the VAMCT and hence degrading the structural integrity. It can be clearly seen that as the

TSR increases, the torque output decreases. For TSR = 0.4, it is clear from the figure that the torque output is negative.

The average torque and power outputs for one complete revolution of the VAMCT are shown in table 1. It is evident that as the TSR increases, the power output decreases. For a TSR of 0.4, the power output is negative which means that the VAMCT is in churning condition and the instead of generating power, the VAMCT is actually consuming power. This condition is of no use as far as power generation from VAMCTs is concerned and hence dictates the upper limit of the TSR under investigation.

Table 1. Average Torque and Power outputs at various TSRs

TSR	Average Torque Output (Nm)	Average Power Output (W)
0.01	72.34	1.03
0.05	61.53	4.39
0.1	49.26	7.03
0.2	25.52	7.29
0.4	-23.21	-13.25

With respect to TSR = 0.01, the average torque output decreases by 15%, 32%, 65% and 132% for TSRs of 0.05, 0.1, 0.2 and 0.4 respectively. However, the power output first increases and then decreases for increasing TSR. From TSR = 0.01 to 0.05, the average power output increases by 326%, from TSR = 0.05 to 0.1 by 60% and from TSR = 0.1 to 0.2 by 3.7% respectively. From TSR = 0.2 to 0.4, the average power output from the VAMCT decreases by 281% indicating that the maximum power output lies between TSR = 0.1 and 0.4.

In order to find out the optimal operating condition for the VAMCT that corresponds to the maximum power output, average torque and power outputs for various TSR values have been plotted together in figure 6. Furthermore, angular speed has also been plotted on a separate scale for comparison. It can be seen that at TSR = 0.315 both the average torque and the average power outputs reaches zero values. At TSR higher than 0.315, both the performance outputs of the VAMCT have negative values showing that the VAMCT is in churning condition. The torque curve shows a constantly decreasing trend w.r.t. TSR whereas the angular speed of the VAMCT keeps on increasing for the same incident flow velocity. However, the average power output first increases and then decreases as TSR increases. The peak of the average power output curve lies at a TSR of 0.17 indicating that the optimal operating condition for the VAMCT at 1 m/sec of incident flow of water and operating at 1 m depth corresponds to TSR of 0.17 where the average power output reaches a value of 7.93 W.

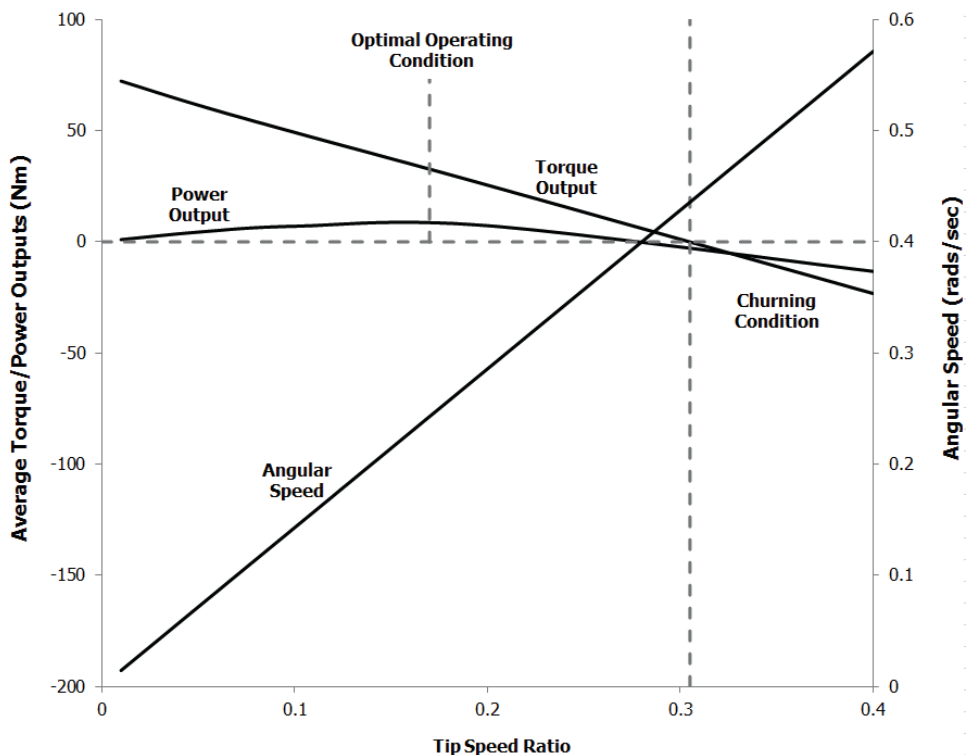


Figure 6. Optimal operating condition of the VAMCT.

4. CONCLUSIONS

A novel design of a Vertical Axis Marine Current Turbine has been numerically simulated for various Tip Speed Ratios. A detailed analysis of the flow field in the vicinity of the VAMCT has shown the presence of high pressure regions at the front end and low pressure regions at the rear end of the VAMCT. Furthermore, it has been shown that as the TSR increases, the torque output from the VAMCT decreases. However, the average output power first increases until it reaches a value of TSR where the average power output from the VAMCT is maximum. This point is regarded as the optimal operating point for the VAMCT. Further increasing the TSR decreases the average torque output from the VAMCT until it reaches the churning condition. Furthermore, it has been shown that Computational Fluid Dynamics based techniques are capable of accurately predicting the flow field in the vicinity of VAMCTs.

5. REFERENCES

- [1] Bai, L. Spence, R.R.G. and Dudziak, G. (2009). "Investigation of the Influence of Array Arrangement and Spacing on Tidal Energy Converter (TEC) Performance using a 3-Dimensional CFD Model", In proceedings of the 8th European Wave and Tidal Energy Conference, Uppsala, Sweden.
- [2] Colley, G. and Mishra, R. (2011). "Computational flow field analysis of a Vertical Axis Wind Turbine", In proceedings of the International Conference on Renewable Energies and Power Quality, Las Palmas, Gran Canaria.
- [3] Dai, Y.M. Gardiner, N. Sutton, R. and Dyson, P.K. (2011). "Hydrodynamic analysis models for the design of Darrieus-type vertical-axis marine current turbines". *Journal of Engineering for the Maritime Environment*, 225, pp. 295-307.
- [4] Li, Y. and Calisal, S.M. (2010). "Numerical analysis of the characteristics of vertical axis tidal current turbines". *Journal of Renewable Energy*, 35, pp. 435-442.
- [5] Paillard, B. Astolfi, J and Hauville, F. (2011). "CFD Simulation and Experimental Validation of a Vertical Axis Turbine: Toward Variable Pitch Cross-Flow Marine Turbine for Maximizing Hydropower Extraction—The SHIVA Project", In proceedings of the 30th International Conference on Ocean, Offshore and Arctic Engineering, Rotterdam, The Netherlands.
- [6] Turnock, S.R. Phillips, A.B., Banks, J. and Lee, R.N. (2011). "Modelling tidal current turbine wakes using a coupled RANS-BEMT approach as a tool for analysing power capture of arrays of turbines". *Journal of Ocean Engineering*, 38, pp. 1300-1307.
- [7] Yang, B. and Lawn, C. (2011). "Fluid dynamic performance of a vertical axis turbine for tidal currents". *Journal of Renewable Energy*, 36, pp. 3355-3366.
- [8] Yang, B. and Shu, X.W. (2012). "Hydro foil optimization and experimental validation in helical vertical axis turbine for power generation from marine current". *Journal of Ocean Engineering*, 42, pp. 35-46.
- [9] Colley, G. and Mishra, R. (2010). "Effect of rotor blade position on Vertical Axis Wind Turbine performance", In proceedings of the International Conference on Renewable Energies and Power Quality (ICREPQ'10) Granada, Spain.
- [10] Park, K.S. Asim, T. and Mishra, R. (2012). "Computational Fluid Dynamics based fault simulations of a Vertical Axis Wind Turbine". *Journal of Physics: Conference Series*, 364.

Application of Particle Swarm Optimization Approach in the Inflationary Inventory Model under Stochastic Conditions

Seyed Mostafa Orand
Department of Industrial
Engineering, Islamic Azad
University, Saveh Branch, Iran
(+98) 2634567682
Smorand00@gmail.com

Abolfazl Mirzazadeh
Assistant Prof.
Department of Industrial
Engineering, Islamic Azad
University, Karaj Branch, Iran
(+98) 912 3054337
A.Mirzazadeh@aut.ac.ir

Farzaneh Ahmadzadeh
Post Doc Researcher,
Luleå University of
Technology, Sweden
(+46)920492106
Farzaneh.Ahmadzadeh@ltu.se

1. ABSTRACT

This paper presents the inventory model considering inflation under non-deterministic situations, which can be used for arbitrary probability density function (pdf), for the inflation rate. We used two solution methods for determining the optimal ordering policy: (1) classical optimization method and (2) particle swarm optimization (PSO) method with comparing the results. The shortages are allowed and fully backlogged. A constant fraction of the on-hand inventory deteriorates per unit time, as soon as the item is received into inventory. The numerical examples are provided to explore the correctness of theoretical results, which are further clarified through a sensitivity analysis on the parameters of the model.

Keywords

Inventory, Inflation, Stochastic, Particle swarm optimization (PSO).

2. INTRODUCTION

Many studies about inventory management systems have been reported. This literature contains analysis of inventory system management using two methods to show the different of classic and modern method. First method, determine the optimal values of the control variables by minimizing the average annual cost using classic numerical method and second is particle swarm optimization.

Since 1975, a series of papers appeared that considered the effects of inflation on inventory system. Buzacott (1975) deal with an economic order quantity model with inflation subject to different types of pricing policies. Misra (1979) developed a discounted cost model and included internal (company) and external (general economy) inflation rates for various costs associated with an inventory system. Sarker and Pan (1994) surveyed the effects of inflation and the time value of money in quantity order with finite replenishment rate. Some efforts were extended to consider variable demand, such as Vrat and Padmanabhan (1990) Datta and Pal (1991), Hariga (1995) Hariga and Ben-Daya (1996) and Chung (2003).

In above cases, it has been implicitly assumed that the rate of inflation is known with certainty. But, inflation enters the inventory picture only because it may have an impact on the future inventory costs yet, and the future rate of inflation is

inherently uncertain and unstable. Horowitz (2000) discussed an Economic Order Quantity (EOQ) model with a normal distribution for the inflation rate and Mirzazadeh and Sarfaraz (1997) presented multiple-items inventory system with a budget constraint and the uniform distribution function for external inflation rate. Finally Mirzazadeh (2007) compare two methods of inventory models under nondeterministic condition but all of studies use classic numerical method to show the results of method and modern method is never use in these conditions. Particle swarm optimization (PSO) algorithm was originally proposed by Kennedy and Eberhart (1995) and many researcher has been use this algorithm for solve non-deterministic polynomial-time hard (NP-Hard) problem. The remainder of this paper is organized as follows. Section 3 provides an overview of the average annual cost model. Section 4 discusses particle swarm optimization method and our improvement setup. In section 5 numerical examples with considering the different methods are provided. Finally, section 6 concludes with a discussion of future research.

3. THE AVERAGE ANNUAL COST MODEL

3.1 Assumptions

The mathematical models in this paper are developed based on the following assumptions:

1. The inflation rate is a random variable with known distribution. The demand rate is known and constant.
2. Shortages are allowed and fully backlogged.
3. The replenishment is instantaneous and the replenishment cycle is the same for each period.
4. Lead time is negligible and the initial inventory level is zero.
5. The time horizon is infinite.
6. A constant fraction of the on-hand inventory deteriorates per unit time, as soon as the item is received into inventory.

3.2 Notations

1. i The inflation rate per unit time
2. $f(i)$ The probability density function (p.d.f.) of i
3. μ The mean value of i
4. σ The standard deviation of i
5. r The interest rate per unit time
6. D The demand rate per unit time

7. θ The constant deterioration rate, where $0 < \theta^2 < 1$
8. c_m The inventory holding cost (for $m=1$) and shortage cost (for $m=2$) per unit per Unit time at time zero
9. c The per unit purchase cost of the item at time zero
10. A The ordering cost per order at time zero
11. T The replenishment time interval
12. k The proportion of time in any given inventory cycle which orders can be filled from the existing stock
13. $EAC(k, T)$ The expected present value of costs
14. N Number of Particle
15. i Iteration
16. $X_j^{(i)}$ Position particle j in i th iteration
17. $V_j^{(i)}$ Velocity particle j in i th iteration
18. $f[X_j^{(i)}]$ Fitness particle j in i th iteration
19. $P_{Best,j}$ Best personal position of particle j in i th iteration
20. G_{Best} Best group position of particles in i th iteration
21. c_1 Cognitive (individual) learning rate
22. c_2 Social (group) learning rate
23. r_1, r_2 Uniformly distributed random numbers in the range 0 and 1
24. θ Coefficient of improvement PSO

3.3 Model formulation

During the time interval $[0, kT]$, the level of inventory $I_1(t_1)$ gradually decreases mainly to meet demands and partly due to deterioration. Hence, the variation of inventory with respect to time can be described by the following differential equation:

$$\frac{dI_1(t_1)}{dt_1} + \theta I_1(t_1) = -D \quad 0 \leq t_1 \leq kT \quad (1)$$

The shortages occur at time kT and accumulated until T before they are backordered. The shortages level be represented by

$$\frac{dI_2(t_2)}{dt_2} = -D \quad 0 \leq t_2 \leq (1-k)T \quad (2)$$

The solution of the above differential equations after applies the boundary conditions $I_1(kT) = 0$ and $I_2(0) = 0$, are

$$I_1(t_1) = -\frac{D}{\theta} \left[e^{-\theta t_1} / e^{-\theta kT} - 1 \right] \quad (3)$$

$$I_2(t_2) = -D t_2 \quad (4)$$

The total cost in this inventory system includes the replenishment cost, purchase cost, holding cost and shortages costs. The problem has been formulated with the average annual cost method. Let $EACR(k, T)$, $EACP(k, T)$, $EACH(k, T)$ and $EACS(k, T)$ denote the expected annual cost of the ordering, purchasing, carrying and shortage, respectively. If n orders are placed for the entire year, expected average replenishment cost per order is

$$E \left[\frac{1}{n} \sum_{j=0}^{n-1} A \left(1 + \frac{j}{n} \right) \right] \quad (5)$$

Where $A[1+(j/n)i]$ is ordering cost for $(j+1)$ -th cycle. Simplifying above equation and considering $T=1/n$ yields

$$EACR(k, T) = E \left[1 + \frac{i(1-T)}{2} \right] A / 2 = \left[1 + \frac{\mu(1-T)}{2} \right] A / T \quad (6)$$

Let PQ , as the total purchase quantity for each cycle where equals to

$$PQ = \int_0^{kT} D e^{\theta t} dt + \int_{kT}^T D dt = \frac{D(e^{\theta kT} - 1)}{\theta} + DT(1-k) \quad (7)$$

Therefore, the average annual cost of purchasing, similarly replenishment cost, is

$$EACP(k, T) = c \left[\frac{D(e^{\theta kT} - 1)}{\theta} + DT(1-k) \right] \left[1 + \frac{\mu(1-T)}{2} \right] / T \quad (8)$$

The average inventory level during the time interval $[0, kT]$ is

$$\bar{I}_{[0, kT]} = \int_0^{kT} \frac{DT(e^{\theta kT} - \theta kT - 1)}{kT} dt = \frac{DT(e^{\theta kT} - \theta kT - 1)}{\theta k} \quad (9)$$

Inventory level is zero during the time interval $[k, T]$. Hence, the average inventory level over each cycle, $[0, T]$, is given by

$$\bar{I}_{[0, T]} = \bar{I}_{[0, kT]} \frac{kT}{T} = \frac{DT(e^{\theta kT} - \theta kT - 1)}{\theta 2} \quad (10)$$

Therefore, the average annual holding cost is

$$EACH(k, T) = c_i \bar{I}_{[0, T]} \left[1 + \frac{\mu(1-T)}{2} \right] \quad (11)$$

The shortages linearly increase during the time interval $[kT, T]$. Therefore, the average shortages level equals to

$$\bar{B}_{[kT, T]} = B_{max} / 2 = DT(1-k) / 2 \quad (12)$$

The average shortages level over each cycle, $[0, T]$, is

$$\bar{B}_{[0, T]} = \bar{B}_{[kT, T]} \frac{T(1-k)}{T} = DT(1-k)^2 / 2 \quad (13)$$

Hence, the average annual shortages cost is

$$EACS(k, T) = c_s \bar{B}_{[0, T]} \left[1 + \frac{\mu(1-T)}{2} \right] \quad (14)$$

So, the expected average annual cost is given by

$$EAC(k, T) = EACR(k, T) + EACP(k, T) + EACH(k, T) + EACS(k, T) \quad (15)$$

4. PARTICLE SWARM OPTIMIZATION

Particle swarm optimization which we call abbreviated as PSO is an evolutionary computation method which based on swarm intelligence and behaviour of a colony or swarm of insects such as ant, bees; a flock of birds or school of fish. This algorithm based on the Adaptive Culture Model and particle swarm which in first time proposed by Kennedy and Eberhart in 1995. Kennedy with help the actual and probably universal behaviours of individuals, which can be summarized in terms of three principles Evaluate, Compare and Imitate In the social society start modelling of the Mathematical model of this algorithm which can you see the performance of it in these following steps.

4.1 PSO Algorithm

Assume the size of the swarm (number of particles) is N . Usually a size of 20 to 30 particles is assumed for the swarm as a compromise. Generate the initial population of X in the range $X^{(l)}$ and $X^{(u)}$ randomly as X_1, X_2, \dots, X_N . Hereafter, **obtain** $f[X_1(0)], f[X_2(0)], \dots, f[X_N(0)]$.

Find the velocities of particles. All particles will be moving to the optimal point with a velocity. Initially, all particle velocities are assumed to be zero. Set the iteration number as $i = 1$.

In the i th iteration, find the following two important parameters used by a typical particle j :

- The historical best value of $X_j(i)$ (coordinates of j th particle in the current iteration i), $P_{best,j}$, with the highest value of the objective function $f[X_j(i)]$, encountered by particle j in all the previous iterations. The historical best value of $X_j(i)$ (coordinates of all particles up to that iteration), G_{best} , with the highest value of the objective function $f[X_j(i)]$, encountered in all the previous iterations by any of the N particles.
- Find the velocity of particle j in the i th iteration as follows:

$$V_j^{(i)} = V_j^{(i-1)} + C_1 r_1 [P_{best,j} - X_j^{(i-1)}] + C_2 r_2 [G_{best} - X_j^{(i-1)}] \quad (16)$$

- Find the position or coordinate of the j th particle in i th iteration as

$$X_j^{(i)} = X_j^{(i-1)} + V_j^{(i)} \quad (17)$$

Where a time step of unity is assumed in the velocity term in Eq. (17). Evaluate the objective function values corresponding to the particles as $F[X_1(i)], F[X_2(i)], \dots, F[X_N(i)]$.

- Check the convergence of the current solution. If the positions of all particles converge to the same set of values, the method is assumed to have converged. If the convergence criterion is not satisfied, step 4 is repeated by updating the iteration number as $i = i + 1$, and by computing the new values of $P_{best,j}$ and G_{best} . The iterative process is continued until all particles converge to the same optimum solution.

4.2 Improvement PSO

It is found that usually the particle velocities build up too fast and the maximum of the objective function is skipped. Hence an inertia term θ is added to reduce the velocity. Usually, the value of θ is assumed to vary linearly from 0.9 to 0.4 as the iterative process progresses. The velocity of the j th particle, with the inertia term, is assumed as (S. Rao, 2009)

$$V_j^{(i)} = \theta V_j^{(i-1)} + C_1 r_1 [P_{best,j} - X_j^{(i-1)}] + C_2 r_2 [G_{best} - X_j^{(i-1)}] \quad (17)$$

5. NUMERICAL EXAMPLE

To illustrate the models described in Section 3, the following numerical example is discussed. Let us choose the following values for the parameters: $D=1000$ units/year, $A=\$60$ /order, $c=\$1$ /unit, $c_1=\$0.2$ /unit/year, $c_2=\$0.6$ /unit/year, $r=\$0.2$ /\$/year and $\theta=0.25$. The inflation rate has the normal distribution function with mean of $\mu=0.12$ and standard deviations of $\sigma=0.06$ and for PSO parameters $V_j^0=0, N=20, C_1=0.9, C_2=0.9, \theta=0.7, r_1, r_2=Random, i=200$. We solve this model with two methods after formulate and computation the results shown in table 1.

Table 1. Average Annual Cost Optimal Solution

Method	k	T	EAC(k,T)*
Classic Numerical Method	0.60336	0.78	1184.06
PSO Algorithm	0.60263	0.7833793	1184.0623

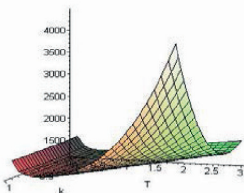


Figure 2. Graphical representation of EAC(k,T) for $0 \leq T \leq 3$

Figure 2 and Figure 3 show changes in the objective functions for average annual cost method with Classic Numerical method and PSO Algorithm respect to k and T.

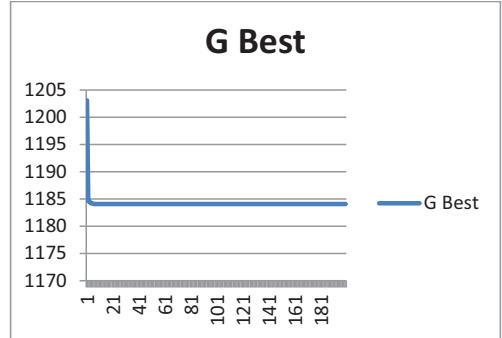


Figure 3. Graphical representation of Gbest

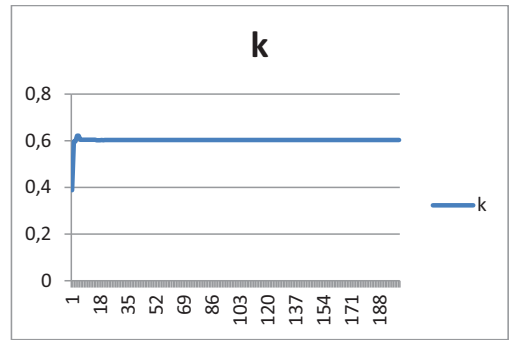


Figure 4. Graphical representation of value k

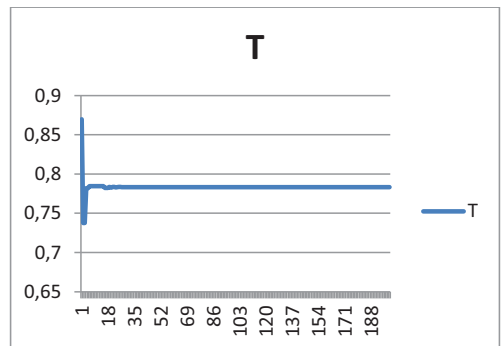


Figure 5. Graphical representation of value T

6. CONCLUSION

In this paper, we prepared an inventory model under stochastic conditions. The two methods have been used to solve the developed model, first the classical method and the second, particle swarm optimization (PSO) method. The objective of the inventory system is the optimal value of Average Annual Cost. The results show that two methods have the same solution, but, the PSO method is very faster. Therefore, the second method will be very useful for this NP-Hard problem.

7. REFERENCE

- [1] Balkhi, Z. T (2004), "On the Optimality of Inventory Models with Deteriorating Items for Demand and On-Hand Inventory Dependent Production Rate." *IMA Journal Management Mathematics* 15, p. 67-86.
- [2] Balkhi, Z. T (2004), "An Optimal Solution of a General Lot Size Inventory Model with Deteriorated and Imperfect Products, Taking into Account Inflation and Time Value of Money." *International Journal of Systems Science*. 35, p. 87-96.
- [3] Buzacott, J. A (1975), "Economic Order Quantities with Inflation." *Operational Research Quarterly*. 26, p. 553-558.
- [4] Chakravarti I. M., Laha R. G. and Roy J (1967), "Handbook of Methods of Applied Statistics." John Wiley & Sons.
- [5] Chang, C. T (2004), "An EOQ Model with Deteriorating Items under Inflation when Supplier Credits Linked to Order Quantity." *International Journal of Production Economics*. 88, p. 307-316.
- [6] Chen J. M (1998), "An Inventory Model for Deteriorating Items with Time-Proportional Demand and Shortages under Inflation and Time Discounting." *International Journal of Production Economics*. 55, p. 21-30.
- [7] Chen, J. M. and C. S. Lin (2002), "An Optimal Replenishment Model for Inventory Items with Normally Distributed Deterioration." *Production Planning and Control*. 13, p. 470-480.
- [8] Chung, K. J (2003), "An Algorithm for an Inventory Model with Inventory-Level-Dependent Demand Rate." *Computers and Operations Research*. 30, p. 1311-1317.
- [9] Chung, K. J. and S. F. Tsai (2001), "Inventory Systems for Deteriorating Items with Shortage and a Linear Trend in Demand-Taking Account of Time Value." *Computers and Operations Research*. 28, p. 915-934.
- [10] Datta, T. K. and A. K. Pal (1991), "Effects of Inflation and Time-Value of Money on an Inventory Model with Linear Time-Dependent Demand Rate and Shortages." *European Journal of operational Research*. 52, p. 326-333.
- [11] Goyal, S. K. and B. C. Giri (2001), "Recent Trends in Modeling of Deteriorating Inventory." *European Journal of Operational Research*. 134, p. 1-16.
- [12] Hadley, G (1964), "A Comparison of Order Quantities Computed Using the Average Annual Cost and the Discounted cost." *Management Science*. 10, p. 472-467.
- [13] Hariga, M. A (1995), "Effects of Inflation and Time-Value of Money on an Inventory Model with Time-Dependent Demand Rate and Shortages." *European Journal of operational Research*. 81, p. 512-520.
- [14] Hariga, M. A. and M. Ben-Daya (1996), "Optimal Time Varying Lot-Sizing Models under Inflationary Conditions." *European Journal of operational Research*. 89, p. 313-325.
- [15] Horowitz, I (2000), "EOQ and Inflation Uncertainty." *International Journal of Production Economics*. 65, p. 217-224.
- [16] Hou, K. L (2006), "An Inventory Model for Deteriorating Items with Stock-Dependent Consumption Rate and Shortages under Inflation and Time Discounting." *European Journal of Operational Research*. 168, p. 463-474.
- [17] Hou, K. L. and L. C. Lin (2004), "Optimal Inventory Model with Stock-Dependent Selling Rate under Maximal Total Present Value of Profits." *Proceeding of the Fourth IASTED International Conference on Modeling, Simulation, and Optimization*. p. 7-12.
- [18] Jaggi C. K., K. K. Aggarwal and S. K. Goel (2006), "Optimal order policy for deteriorating items with inflation induced demand." *International Journal of Production Economics*. 103, p. 707-714.
- [19] Jeffrey, A (1995), "Handbook of mathematical formulas and Integrals." London, Academic Press.
- [20] Liao, H. C. and Y. K. Chen (2003) "Optimal Payment Time for Retailer's Inventory System." *International Journal of Systems Science*. 34, p. 245-253.
- [21] Lo S. T., H. M. Wee, W. C. Huang (2007), "An integrated production-inventory model with imperfect production processes and Weibull distribution deterioration under inflation." *International Journal of Production Economics*. 106., p. 248-260.
- [22] Maiti A.K., M.K. Maiti and M. Maiti (2006), "Two storage inventory model with random planning horizon." *Applied Mathematics and Computation*. 183, p. 1084-1097.
- [23] Mirzazadeh, A. and A. R. Sarfaraz (1997), "Constrained Multiple Items Optimal Order Policy under Stochastic Inflationary Conditions." *second Annual International Conference on Industrial Engineering Application and Practice (USA, San Diego)*. p. 725-730.
- [24] Misra, R. B (1979), "A Note on Optimal Inventory Management under Inflation." *Naval Research Logistics Quarterly*. 26, p. 161-165.
- [25] Moon, I., B. C. Giri and B. Ko (2005), "Economic Order Quantity Models for Ameliorating/Deteriorating Items under Inflation and Time Discounting." *European Journal of Operational Research*. 162, p. 773-785.
- [26] Sarker, B. R., A. M. M. Jamal and S. Wang (2000), "Supply Chain Models for Perishable Products under Inflation and Permissible Delay in Payments." *Computers and Operations Research*. 27, p. 59-75.
- [27] Sarker, B. R. and H. Pan (1994), "Effects of Inflation and the Time Value of Money on Order Quantity and Allowable Shortage." *International Journal of Production Economics*. 34, p. 65-72.
- [28] Spiegel (1968), M. R., "Mathematical Handbook of formulas and tables." McGraw-Hill.
- [29] Su, C. T., L. I. Tong and H. C. Liao (1996), "An Inventory Model under Inflation for Stock Dependent Demand Rate and Exponential Decay." *Operations Research*. 33, p. 71-82.
- [30] Vrat, P. and G. Padmanabhan (1990), "An Inventory Model under Inflation for Stock Dependent Consumption Rate Items." *Engineering Costs and Production Economics*. 19, p. 379-383.
- [31] Wee, H. M. and S. T. Law (1999), "Economic Production Lot Size for Deteriorating Items Taking Account of the

- Time Value of Money.” *Computers and Operations Research*. 26, p. 545-558.
- [32] Wee, H. M. and S. T. Law (2001), “Replenishment and Pricing Policy for Deteriorating Items Taking into Account the Time-Value of Money.” *International Journal of Production Economics*. 71, p. 213-220.
- [33] Yang, H. L (2004), “Two-Warehouse Inventory Models for Deteriorating Items with Shortages under Inflation.” *European Journal of Operational Research*. 157, 2004. p. 344-356.
- [34] Yang, H. L (2006), “Two-warehouse partial backlogging inventory models for deteriorating items under inflation.” *International Journal of Production Economics*. 103, p. 362–370.
- [35] Yang, Teng and Chern (2001), “Deterministic Inventory Lot-Size Models under Inflation with Shortages and Deterioration for Fluctuating Demand.” *Naval Research Logistics*. 48, p. 144-158.
- [36] S. Rao, S. (2009). *Engineering optimization: Theory and practice*. In S. S. Rao, *Engineering optimization: Theory and practice* (Vol. 1, pp. 708-714). Hoboken, New Jersey, USA: John Wiley and Sons.
- [37] Mirzazadeh, A. (2007). *Effects of Uncertain Inflationary Conditions on Inventory Models Using the Average Annual Cost and the Discounted Cost*. *International Conference on Operations and Quantitative Management* (pp. 658-666). Bangkok: AIMS.
- [38] Kennedy, J., Eberhart, R., & Shi, Y. (2001). *Swarm Intelligence*. In J. Kennedy, R. C. Eberhart, & Y. Shi, *Swarm Intelligence* (Vol. 1, pp. 287-314). San Francisco, CA, USA: Morgan Kaufmann.

Multi-objective reliability allocation problem

Jalal Safari
Assistant Prof., Dep. of Industrial
Engineering, Karaj Branch, Islamic
Azad University, Karaj, Iran.
Tel.: + 98 263 4182333
safari@kiau.ac.ir

Farzaneh Ahmadzadeh
Postdoctoral researcher, Dep. of
Operation and Maintenance Eng.
LTU University, Lulea, Sweden.
Tel.: + 46 722 460950
farzaneh.ahmadzadeh@ltu.se

ABSTRACT

This paper proposes a Non-dominated Sorting Genetic Algorithm (NSGA-II) to solve the developed model for multi-objective max-min redundancy allocation problems (RAPs). In this paper, the model for the multi-objective max-min RAPs (MOMRAPs), already presented by others, is developed. To develop this model the pre-assumption that the redundancy strategies of all subsystems are active and predetermined, is removed. Selecting the best redundancy strategy, either active or cold standby, for each subsystem can improve the reliability of systems. Also, the proposed model can be used for any system structure such as series-parallel system, complicated system, and the like. This problem belongs to the NP-hard class. Therefore, this paper presents an NSGA-II to find the non-dominated solutions for the given problem. Finally, the advantages of the presented multi-objective model and the proposed algorithm are illustrated by solving test problems taken from the literature and the efficiency of the proposed method is discussed.

Keywords

Reliability · Optimization · Multi-objective redundancy allocation problem · Non-dominate sorting genetic algorithm-II.

1. INTRODUCTION

The RAP involves the simultaneous selection of components and a system-level design configuration that can collectively meet all design constraints in order to optimize some objective functions, such as system reliability and/or cost. In this problem, there are several types of different components with different levels of cost, reliability, weight, and other characteristics. The majority of the solution methods for the general single objective RAP (SORAP) assume that the redundancy strategy for each subsystem is predetermined and fixed. In general, active redundancy has received more attention than other redundancy strategies. However, in practice both active and cold-standby redundancies may be used within a particular system design. Coit (2003) presented an optimal solution to the SORAP when there are some subsystems using active redundancy, cold-standby redundancy, or selecting the best redundancy strategy. This problem belongs to the NP-hard class of problems. Tavakkoli-Moghaddam *et al.* (2008) and Safari and Tavakkoli-Moghaddam (2010) applied a Genetic Algorithm (GA) and a Memetic Algorithm (MA) to solve this problem, respectively. Tavakkoli-Moghaddam and Safari (2007) developed the mathematical model for the SORAP which was presented by Coit (2003). They formulated a new mathematical model for the SORAP with mixing component when there are some subsystems using active redundancy, cold-

standby redundancy, or selecting the best redundancy strategy. The simple SORAP is known to be an NP-hard problem (Chern, 1992), and has been solved by using a number of optimization approaches for different formulations, as summarized in Gen and Yun (2006).

Most SORAP studies have concentrated on maximizing system reliability or minimizing system cost. However, in practice, a decision maker (DM) or a system designer has to consider multiple objectives simultaneously. In addition to reliability, cost, weight and volume are also important characteristics considered in the system level. Therefore, multiple objective functions become an essential aspect in the reliability design of engineering systems. It is unusual for engineering design problems to have multiple competing objectives and multiple prospective solutions and one cannot identify a single solution that simultaneously optimizes each objective. When attempting to improve an objective further, other objectives suffer as a result. A tentative solution is called non-dominated or Pareto optimal if it cannot be eliminated from consideration by replacing it with another solution which improves an objective without worsening another one. Finding such non-dominated solutions is the goal of setting up and solving a multi-objective optimization problem. There has been a growing interest in the Multi-objective RAP (MORAP). Liang and Lo (2010) developed a variable neighborhood search algorithm to solve the MORAP. They considered three types of MORAP to maximize system reliability and minimize system cost or system weight. Lins and Drogue (2011) proposed a multi-objective genetic algorithm coupled with discrete event simulation to solve the MORAP in systems subject to imperfect repairs. They modeled the failure-repair processes of system components by Generalized Renewal Processes. Their presented methodology provides a set of compromise solutions that incorporate not only system configurations, but also the number of maintenance teams. Khalili-Damghani and Amiri (2012) proposed a procedure based on efficient epsilon-constraint method and data envelopment analysis (DEA) to solve a binary-state MORAP.

Since time-to-failure of the system is dictated by the minimum subsystem time-to-failure, a logical design strategy is to increase the minimum subsystem reliability as high as possible, given constraints on the system. For some system design problems, a preferred design objective may be to maximize the minimum subsystem reliability. Additionally, the max-min formulation can serve as a useful and efficient surrogate for optimization problems to maximize system reliability. This strategy can be vital especially for security systems like radars, satellites, detectors internet banking, etc. Ramirez-marquez *et al.* (2003) formulated the SORAP with the objective of maximizing the minimum subsystem reliability for a series-parallel system and use integer

programming methods to determine system design configurations that allow mixing of functionally equivalent component types within a subsystem. Soyulu and Ulusoy (2011) consider a MORAP to maximize the minimum subsystem reliability, while minimizing the overall system cost. They found the Pareto solutions of this problem by the augmented ϵ -constraint approach for small and moderate sized instances. They assumed that the type of redundancy strategy for each subsystem is active. In practice both active and cold-standby redundancies may be used within a particular system design.

In this paper, the MOMRAP which was presented by Soyulu and Ulusoy (2011) is developed and the choice of redundancy strategy for each subsystem becomes an additional decision variable. Selecting the best redundancy strategy (active or cold standby) for each subsystem can improve the reliability of system. Then, in this paper, the MOMRAP is formulated when there are some subsystems using active redundancy, cold-standby redundancy, or selecting the best redundancy strategy. As it was mentioned, Chern (1992) showed that even a simple SORAP in series systems with linear constraints is an NP-hard problem, prompting researchers to develop meta-heuristic methods to achieve near-optimal solutions of acceptable quality in reasonable computational time. Therefore, a well-known MOEA named NSGA-II is utilized to solve this problem.

2. Problem formulation

In this paper, the MORAP is considered to maximize the minimum subsystem reliability, while minimizing the overall system cost. The notations and the proposed formulation are as follows:

Notations

i	Index of subsystem ($i = 1, 2, \dots, s$)
j	Index of component type
s	Number of subsystems
m_i	Number of available component choices for subsystem i
m	Upper bound of m_i , ($m_i \leq m \quad \forall i$)
n_{ij}	Number of components of type j used in subsystem i
n_i	Set of n_{ij} (n_{i1}, n_{i2}, \dots)
$n_{\max, ij}$	Upper bound for n_{ij} , ($n_{ij} \leq n_{\max, ij} \quad \forall ij$)
p_i	Number of the allocated component in subsystem i , ($p_i = \sum_{j=1}^{m_i} n_{ij} \quad \forall i$)
k_i	Number of failed components of type j in subsystem i
l	Number of allocated component types
z_{il}	Index of component choices for subsystem i with l number of allocated component types, $z_{il} \in \{1, 2, \dots, m_i\}$
z_i	Set of z_{il} , (z_{i1}, z_{i2}, \dots)
ARS_i	Assigned redundancy strategy for subsystem i , ($i = 1, 2, \dots, s$)
$R_i(t; n_i, ARS_i)$	Reliability of subsystem i
t	Mission time
$r_{ij}(t)$	Reliability at time t for the j^{th} available component for subsystem i
$\rho_i(t)$	Failure-detection/switching reliability at time t

$f_{ij}^{k_j}(t)$ Probability density function (pdf) of the k_j^{th} failure times of components of type j for subsystem i at time t

c_{ij}, w_{ij} Cost and weight for the j^{th} available component for the subsystem i

W System-level constraint limit for weight

C System-level constraint limit for cost

ff_1 Fitness function of the first objective function

ff_2 Fitness function of the second objective function

According to the notations, proposed mathematical model for the MOMRAP is shown as follows:

$$\text{Max Min } R_i(t; n_i, ARS) \quad (1)$$

$$\text{Min } C = \sum_{i=1}^s \sum_{j=1}^{m_i} c_{ij} n_{ij} \quad (2)$$

s.t.

$$\sum_{i=1}^s \sum_{j=1}^{m_i} w_{ij} n_{ij} \leq W \quad (3)$$

$$n_{ij} \leq n_{\max, ij} \quad (4)$$

With respect to Eq. (1) and (2), the objective is to determine the redundancy strategy, combination of components, and quantity of components in each subsystem to maximum the minimum subsystem reliability and minimize the corresponding designing cost. Constraints given in Eqs. (3) and (4) consider the available weight and maximal number of components hold for subsystem i , respectively. This model is proposed by Soyulu and Ulusoy (2011). As it mentioned, Tavakkoli-Moghaddam and Safari (2007) presented the mathematical model, for the SORAP to maximize the reliability of system when there are some subsystems using active redundancy, cold-standby redundancy, or selecting the best redundancy strategy and Safari and Tavakkoli-Moghaddam (2009) show the way of system reliability calculation for the problem. In this paper we use the mathematical model which was presented by Safari and Tavakkoli-Moghaddam (2009) to calculate the reliability of subsystems.

3. Solution Method: Non-dominated Sorting Genetic Algorithm (NSGA-II)

The NSGA uses a ranking method that emphasizes on the good solution points and tries to maintain a population of such points throughout the procedure. NSGA maintains diversity in its population by a sharing method. This eliminates focusing on certain regions of the solution space, and explores different regions in the Pareto front. This algorithm is highly efficient in obtaining good Pareto optimal fronts for any number of objectives and can handle any number of constraints. NSGA-II is a well known and extensively used algorithm based on its predecessor NSGA. It was formulated by Deb *et al.* (2002) as a fast and very efficient NSGA, which incorporates an elitist archive and a rule for adaptation assignment that takes into account both the rank and the distance of each solution regarding others. In this paper, the a variant NSGA-II will employ, since there have been many investigations ensuring that this solution method can often converge to Pareto-optimal set and the obtained solutions can often spread well over the Pareto-optimal set. In this section, on the basis of the system model, appropriate coding scheme and operators for NSGA-II is chosen to adapt this algorithm to MOMRAP when the redundancy strategy can be chosen for individual subsystems.

3.1 Solution Encoding

Each possible solution to the given problem is a collection of the redundancy strategy and the number of allocated components of different types for each subsystem. The solution encoding for this problem is a $s \times (m+1)$ matrix. The first column of the matrix represents allocated redundancy strategy in each subsystem, and the other columns represent the number of the allocated components of different types in each subsystem. The subsystem representations are then placed into adjacent rows to complete the matrix representation.

3.2 Initial Population

The initial feasible population is determined randomly by selecting p solutions, wherein N is a population of size 100. This population size is used for a lot of researches like Deb *et al.* (2002). Some problems have very large solution spaces (i.e. many variables, with a large range of permissible values for those variables). In these cases, the population size must be selected no less than 100.

3.3 Fitness function

One of the main issues in evolutionary computation is how to guide the search towards the feasible region in the presence of constraints. The existing approaches can be classified in the repair techniques, separation techniques, penalization techniques, and hybrid techniques groups (Safari, 2012). In this paper, a penalization technique is utilized. Therefore, the fitness function is defined as the summation of the objective and a dynamic penalty function determined by the relative degree of infeasibility. To provide an efficient search through the infeasible region but to ensure that the final best solution is feasible, the following fitness functions are proposed.

$$ff_1 = \begin{cases} \text{Max min } R_i(t; n, ARS) & \text{if } \sum_{i=1}^s w_{iz_i} n_i \leq W \\ \text{Max min } R_i(t; n, ARS) \times \left(\frac{\sum_{i=1}^s w_{iz_i} n_i - W}{W} \right) & \text{Otherwise} \end{cases} \quad (5)$$

$$ff_2 = \begin{cases} C & \text{if } \sum_{i=1}^s w_{iz_i} n_i \leq W \\ C + (\sum_{i=1}^s w_{iz_i} n_i - W) & \text{if } \sum_{i=1}^s w_{iz_i} n_i > W \end{cases} \quad (6)$$

3.4 Crossover Breeding Operator

The proposed crossover breeding operator is applied with crossover probability (rate) of 0.9. This crossover rate is used for a lot of researches like Deb *et al.* (2002). In this paper, it is found this rate to be best according to our experimental results. One offspring is produced by a uniform crossover breeding operator. This operator first generates a random crossover mask and then exchanges the respective subsystems between parents according to the mask and the gen of parents will be exchange if the respective bit in mask matrix is 1. The crossover mask is simply a binary $s \times 1$ matrix. Another offspring is produced by the modified uniform crossover operator. For each candidate solution, this operator increases the min-max subsystem reliability by making the reliability of subsystems closer to each other and the crossover mask is not generated randomly. The crossover mask is intelligently generated as follows:

- A $s \times 1$ zero crossover mask matrix is generated. All bits of this matrix are zero.

- Subsystems with the lowest and highest reliability among the candidate solution are determined.
- All bits of these two selected subsystems in $s \times 1$ zero crossover mask matrix are changed to 1.

3.5 General Mutation Operator

The mutation operator performs random perturbations to the selected solutions. Each value within the solution matrix is altered at random with a predefined mutation rate. The mutation rate of $p_m = \frac{1}{d}$ (where d is the number of decision variables) is used. This rate has also been found to work well for most NSGA II solutions (Deb *et al.*, 2002).

3.6 Max-min Mutation Operator

Subsystem with lowest reliability has a bad effect on the system reliability. Also, subsystem with highest reliability has no great effect on the system reliability. For improving the reliability of subsystem with lowest reliability, some components are added to the subsystem by mutation. For preventing the creation of infeasible solutions, we remove some components from the subsystem with highest reliability. In other words, by making the reliability of subsystems closer to each other, we want to increase max-min subsystem reliability. Based on the description, the max-min mutation is constructed. For each candidate solution, this operator selects subsystems with the highest and lowest reliability among the subsystems and then randomly mutates the decision variables of each selected subsystem.

3.7 Selection and Stopping Condition

After operator breeding, the N non-dominated solutions among the previous generation and the new children are retained to form the next generation based on step 3 which was mentioned in pseudo-code of NSGA-II. The NSGA-II is terminated after a pre-selected number of generations. A reasonable number of generations is 250.

4. A numerical example

To evaluate the performance of the proposed NSGA-II, a typical example is first solved to determine the Pareto optimal set. This example is a modified version of the problem studied by Fyffe *et al.* (1968). A series-parallel system is connected by 14 parallel subsystems, and each subsystem has three or four components of choice. Time to failure distribution for all components is negative exponential. The purchase cost of each element is assumed to be an increasing function of its expected lifetime. In other words, it is possible to increase the expected lifetime of each element by paying higher purchase cost. Component cost, weight, and the negative exponential distribution parameters are given in Table 1. We are interested in the optimal allocation. The objective is to maximize the minimum subsystem reliability at $t = 100$ hours and minimize system cost, given the constraint for the system weight ($W \leq 170$). In other words, the problem is to determine how much should be invested in the reliability of the elements so that the total costs of the system (initial purchase cost) is minimized and the minimum subsystem reliability is maximized under some constraints. For each subsystem, active or cold-standby redundancy can be used, and the reliability of a switch at 100 hours is 0.99 for all subsystems. Selecting the best redundancy strategy for each subsystem is a decision variable considered in this paper. There is a limit on the total number of redundant

components that can be used. The maximum number of allocated components is 6 (i.e., $\sum_{j=1}^4 n_{ij} \leq 6$) within any subsystem.

The proposed NSGA-II is used to solve this problem. Because of the stochastic nature of the proposed NSGA II, 7 experiments are performed for 250 generation and the non-dominated solutions among these runs are selected and reported in table 2 and figure 2. Among these runs, 21 non-dominated solutions are obtained from the proposed variant of NSGA-II. After finding the Pareto optimal set of non-dominated solutions, DM can select a best-compromised solution among all non-dominated solutions. Figure

4 shows the solutions of the initial population which is created randomly and the solutions of the last population which is prepared by the proposed NSGA II in one run. Each spot in this figure shows the solution obtained by the proposed algorithm. This figure shows the improvements of solutions and the efficiency of the proposed method. Table 3 presents two of the non-dominated solutions (extreme points) obtained by the proposed NSGA-II and the optimal solutions which are obtained by Soylu and Ulusoy (2011). Note that the results of the proposed solution method are near optimal solution. These approve the efficiency of the proposed solution method.

Table 1: Parameter setting for the given problem.

i	Choice 1 (j = 1)			Choice 2 (j = 2)			Choice 3 (j = 3)			Choice 4 (j = 4)		
	λ_{ij}	c_{ij}	w_{ij}	λ_{ij}	c_{ij}	w_{ij}	λ_{ij}	c_{ij}	w_{ij}	λ_{ij}	c_{ij}	w_{ij}
1	0.001053605	1	3	0.000725707	1	4	0.000943107	2	2	0.000512933	2	5
2	0.000512933	2	8	0.000618754	1	10	0.000725707	1	9	-	-	-
3	0.001625189	2	7	0.001053605	3	5	0.001392621	1	6	0.000833816	4	4
4	0.001863296	3	5	0.001392621	4	6	0.001625189	5	4	-	-	-
5	0.000618754	2	4	0.000725707	2	3	0.000512933	3	5	-	-	-
6	0.000100503	3	5	0.000202027	3	4	0.000304592	2	5	0.00040822	2	4
7	0.000943107	4	7	0.000833816	4	8	0.000618754	5	9	-	-	-
8	0.00210721	3	4	0.001053605	5	7	0.000943107	6	6	-	-	-
9	0.000304592	2	8	0.000100503	3	9	0.00040822	4	7	0.000943107	3	8
10	0.001863296	4	6	0.001625189	4	5	0.001053605	5	6	-	-	-
11	0.000618754	3	5	0.000512933	4	6	0.00040822	5	6	-	-	-
12	0.002357223	2	4	0.001984509	3	5	0.001625189	4	6	0.001053605	5	7
13	0.000202027	2	5	0.000100503	3	5	0.000304592	2	6	-	-	-
14	0.001053605	4	6	0.000833816	4	7	0.000512933	5	6	0.000100503	6	9

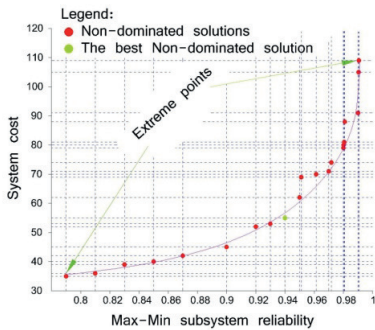


Figure 1. The non-dominated solutions.

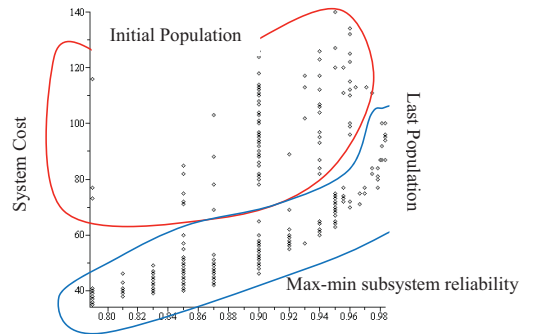


Figure 2. The initial and last population of one run.

Table 2: The non-dominated solutions which are identified by the proposed NSGA II.

Answer number	Max-min subsystem reliability	System cost	System weight	Run number
1	0.79	35	78	26
2	0.81	36	83	39
3	0.83	39	87	35
4	0.85	40	88	35
5	0.87	42	87	26
6	0.9	45	99	39
7	0.92	52	114	26
8	0.93	53	109	39
9	0.94	55	113	16
10	0.95	62	132	26
11	0.9559	69	145	12
12	0.9665	70	147	44
13	0.97	71	143	26
14	0.9785	74	150	26
15	0.9805	79	151	44
16	0.981	80	163	26
17	0.9831	81	161	26
18	0.985	88	166	16
19	0.99	91	164	45
20	0.9919	105	166	12
21	0.9931	109	170	45

Table 3: Comparison between computational results of the proposed NSGA-II and the optimal solution

Non-dominated solution	Proposed NSGA II				Optimal solution (Soylu and Ulusoy, 2011)	
	Max-min subsystem reliability	C	W	Address at table 2	Max-min subsystem reliability	C
Lowest extreme point	0.79	35	78	First row	0.79	34
Highest extreme point	0.9931	109	170	Last row	0.9958	115

5. Conclusion

In this paper, a bi-objective max–min RAP has been studied for different system structures, where either active or cold-standby redundancy can be selected for individual subsystems. The choice of redundancy strategies for each subsystem is much more realistic and provides reliability engineers with a better tool for designing systems. We have formulated the problem as a multi-objective nonlinear integer programming model subject to a number of physical constraints. This problem belongs to the NP-

hard class. This motivates the use of meta-heuristic methods for solving such a hard and complex problem. Therefore, a variant of NSGA-II has been proposed for finding non-dominated solutions to this problem. Finally, the results of the solution method were compared with the optimal solution. This comparison approved the efficiency of the proposed method.

6. ACKNOWLEDGMENTS

Our thanks to KIAU for intellectual and financial support.

7. REFERENCES

[1] Chern, M. S. (1992). "On the computational complexity of reliability redundancy allocation in a series system". *Operation Research Letters*, 11, pp. 309-315.

[2] Coit, D. W. (2003). "Maximization of system reliability with a choice of redundancy strategies". *IIE Transactions*, 35(6), pp. 535-544.

[3] Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T. (2002). "A fast and elitist multi-objective genetic algorithm: NSGA-II". *IEEE Transactions on evolutionary computation*, 6(2), pp. 182-196.

[4] Fyffe, D. E., Hines, W. W. and Lee, N. K. (1968). "System reliability allocation and a computational algorithm". *IEEE Transactions on Reliability*, 17, pp. 64-69.

[5] Gen, M. and Yun, Y. S. (2006). "Soft computing approach for reliability optimization: State-of-the-art survey". *Reliability Engineering and System Safety*, 91(9), pp. 1008-1026.

[6] Khalili-Damghani, K. and Amiri, M. (2012). "Solving binary-state multi-objective reliability redundancy allocation series-parallel problem using efficient epsilon-constraint, multi-start partial bound enumeration algorithm, and DEA". *Reliability Engineering and System Safety*, 103, pp. 35-44.

[7] Liang, Y-C. and Lo, Y-C. (2010). "Multi-objective redundancy allocation optimization using a variable neighborhood search algorithm". *Journal of Heuristics*, 16, pp. 511-535.

[8] Lins, I. D. and Droguett, E. L. (2011). "Redundancy allocation problems considering systems with imperfect repairs using multi-objective genetic algorithms and discrete event simulation". *Simulation Modeling Practice and Theory*, 19, pp. 362-381.

[9] Ramirez-Marquez, J. E., Coit, D. W. and Konak, A. (2004). "Redundancy allocation for series-parallel systems using a max-min approach". *IIE Transactions*, 36, pp. 891-898.

[10] Safari, J. (2012). "Multi-objective reliability optimization of series-parallel systems with a choice of redundancy strategies". *Reliability Engineering and System Safety*, 108, pp. 10-20.

[11] Safari, J. and Tavakkoli-Moghaddam, R. (2009). "A Memetic algorithm for a redundancy allocation problem". Proceeding of the 5th International Conference on Industrial Engineering, Tehran of Iran.

[12] Safari, J. and Tavakkoli-Moghaddam, R. (2010). "A redundancy allocation problem with the choice of redundancy strategies by a memetic algorithm". *Journal of Industrial Engineering International*, 6(11), pp. 6-16.

[13] Soylu, B. and Ulusoy, S. K. (2011). "A preference ordered classification for a multi-objective max–min redundancy allocation problem". *Computers and Operations Research*, doi:10.1016/j.cor.2011.02.024.

- [14] Tavakkoli-Moghaddam, R. and Safari, J. (2007). "A new mathematical model for a redundancy allocation problem with mixing components redundant and choice of redundancy strategies". *Applied Mathematical Sciences*, 45(1), pp. 2221-2230.
- [15] Tavakkoli-Moghaddam, R., Safari, J. and Sassani, F. (2008). "Reliability optimization of series-parallel systems with a choice of redundancy strategies using a genetic algorithm". *Reliability Engineering and System Safety*, 93, pp. 550-556.

On the Integration of Wear Model into Dynamic Analysis for Rolling Element Bearing

Idriss El-Thalji
Industrial system,

VTT Technical research centre of Finland
Espoo, 02044, Finland
idriss.el-thalji@vtt.fi

Erkki Jantunen
Industrial system,

VTT Technical research centre of Finland
Espoo, 02044, Finland
erkki.jantunen@vtt.fi

ABSTRACT

Wear modelling is one of the greatest issues for wear fault monitoring and it belongs to condition-based maintenance practices. The early prediction of the fault might provide a more cost effective way for maintenance. It might have the potential to actively influence the design procedures of materials and tribology contact. However, it is the core to gain better understanding of the deterioration process and to achieve better use of Predictive Health Monitoring (PHM) procedures. There are a number of dynamic analyses of spalled components, single-wear prediction models and integrated models considering dynamic analysis and wear predictions. However, rolling element bearing requires more sophisticated models to study wear fault taking into account multi-wear mechanisms and the sliding rolling impact phenomenon. The paper starts with a review of the main contribution within wear prediction modelling and system dynamic analysis. Based on the relevant contributions, a numerical dynamic wear analysis for rolling element bearing is proposed. In the paper, a comparative study was performed based on the empirical findings of real rolling bearing applications and the collected modelling contributions. The paper highlights a number of practical considerations that are needed to be modelled within the whole integrated prediction of wear process. Moreover, the integrated analysis model relies highly on the incremental feedback relationship between basic modules. A wear modelling tool has potential to provide more accurate wear prediction and better support for diagnosis and prognosis purposes. This might lead to reduction of the overall operational losses and their associated costs once it is accurate.

Keywords

Wear, Modelling, Sliding, Rolling, Impact contact, Multi-wear mechanism, Rolling bearing

1. INTRODUCTION

Modelling the wear process in rolling element bearing (REB) requires a number of analyses to be considered. First, a system dynamic analysis is required since it provides fundamental outcomes as forces, position, velocity and acceleration. Second, the wear analysis provides updated wear amount and surface characteristics. Third, an integration module is required in order to generate a feedback link between the system dynamic analysis and wear analysis to transfer the updated data between them. The updated forces and position are required to perform

the wear analysis based on the data that are provided by the system dynamic analysis, in particular, to estimate the contact pressure and the slip distance. In the same manner, the outcomes from the wear analysis should iteratively be transferred into the system dynamic module based on the updated surface and contact characteristics (i.e. clearances). Fourth, a signal processing module is needed to analyse the generated signals by the system dynamic modules in order to decide whether the machine behaviour is healthy or faulty. For example, Watson et al [1] had developed an integrated model for high power clutch system and Mukras et al, [2] for a slider-crank mechanism. Such described models set up a pilot step to integrate all these modules into one integrated numerical model for the rolling element bearing.

For the rolling element bearing applications, there are a number of dynamic models and signal analysis methods. The dynamic models of rolling element bearing were first developed by Palmgren [3] and Harris [4]. However, the non-linearity and the time varying characteristics were not addressed. Therefore, Gupta [5] completed the first dynamic model of REB and later Fukata et al. [6] presented a comprehensive non-linear and time variant model. More advanced issues of time variant characteristics and non-linearity were raised and studied by several authors. For example, Wijnat et al. [7] reviewed the studies concerning the effect of EHL on the dynamics of REB. Tiwari, et al. in [8] and [9] studied the effect of the ball bearing clearance on the dynamic response of a rigid rotor. Sapanen & Mikkola [10] reviewed different dynamic models with discussion of the effect of waviness, EHL, and localised faults and clearance effect.

In the finite level of the contact, Kiral & Karagülle [11] presented a defect detection method using the finite element vibration analysis for rolling element bearings with single and multiple defects. The vibration signal function includes impulses produced by the fault, modulation effect due to non-uniform load distribution, bearing induced vibrations, and machinery induced vibrations and the noise which is encountered in any measurement system. Sapanen & Mikkola [10] implemented the proposed ball bearing model using a commercial multi-body system software application (MSC.ADAMS). The effect of the diametrical clearance of the bearing on the natural frequencies and vibration response of the rotor-bearing system was studied. The diametrical clearance was found to have a significant effect on the level of vibration

as well as on the natural frequencies. The low-order waviness, also known as the out-of-roundness, is found to generate vibration at the frequencies of the waviness order multiplied by the rotational speed. On the other hand, the waviness orders close to the number of balls in the bearing are found to generate vibrations at the ball passage inner ring and the ball passage outer ring frequencies. Localized defects in the inner and outer ring are found to generate vibrations at the bearing defect frequencies.

Later, a number of combined FEM and multi-body Dynamic models have been presented. Endo [12] developed a 16-degrees-of-freedom (DOF) model of a gearbox in order to simulate spall and cracks in the gear teeth. First, he utilised a FEM model to simulate the variation of the mesh stiffness for two types of faults under varying static load conditions. Then the model was integrated into the lumped parameter dynamic model. The study obtained the dynamic transmission error and the acceleration responses under different loads and speeds. Sawalhi & Randall [13] developed a 34-DOF model of a gearbox in order to simulate spall and cracks in the REB based on Endo's model of 16-DOF. This model includes extra 18-DOF due to considering a five-DOF bearing model and considering a translational DOF both along the line of action and perpendicular to it. Massi et al. [14] studied the wear that is resulting from false brinelling at the contact surfaces between the balls and races of the bearings.

These multi-body dynamic models were highly linked to the vibration and signal analysis. Therefore, they modelled a detailed fault/defect and later introduced it into the whole dynamic model. However, it is noted that these models try to detect the defect within the generated vibration signals and not the incremental deterioration process of the contact surface. For example, it is a relevant issue to model the development of the spall defect size and its deterioration process (i.e. large spall defect sizes) using a specific tracking procedure as performed by Sawalhi et al. [15]. In order to do that, an incremental numerical procedure should be integrated into the multi-body dynamic system that needs a procedure to update the contact information continuously. This means that the applied force and conditions into contact analysis should be updated during every iteration round of the dynamic analysis. It is worth to develop a similar procedure for rolling element bearings on the basis of Watson et al. [1] and Mukras et al. [2]. However, a number of issues should be considered. First, the procedure should be capable to manage multi-wear mechanisms. Second, it should be capable to involve the influences of having sliding, rolling, and impact conditions. Third, it should be capable to link the required inputs into the vibration signal analysis and link the outputs back to the other analysis methods.

2. MODELLING OF ROLLING BEARING DYNAMICS

The dynamic models are always enhanced to consider and represent the real physical properties of rolling element bearing i.e. stiffness, damping, dimensions, Hertzian force/deformation, clearances, Elasto-hydrodynamic lubrication, surface roughness, waviness, etc. There are a number of considerations and delimitations that have been used in the system dynamic models.

First, the mechanism modelling and the degrees of freedom are represented in the equations of motion of rolling element bearing. For example, the inertia, the restoring forces, the damping forces and the applied force acting on the inner race are considered. Purohit & Purohit [16] represented two equations of motion for the ball bearings, where they considered the above mentioned aspects. However, the equations of motion require the calculation of a number of parameters such as contact stiffness, restoring force, etc.

Second, the contact stiffness can be estimated according to the Harris model in [4]. The model is based on the curvature differences and the geometry of contacting bodies.

Third, the restoring forces for each rolling element are considered in the equation of motion based on the local Hertzian contact force and the deflection relationship. In this case, deflection depends on radial deflection after subtracting the internal radial clearance. Moreover, the total restoring force is the sum of the restoring force from each rolling element. The number of rolling elements has also direct relationship with the system stiffness. Increasing the number of rolling elements means increasing the number of supporting points for the shaft, therefore, the number of rolling elements increases the system stiffness and reduces the vibration amplitude as described in [16].

Fourth, the elastic deformation should be considered due to its importance in estimating the contact forces that are acting on the rolling elements. Thus, the total elastic deformation is calculated by considering the distance between the race surfaces along the line of contact and the lubricant film thicknesses between the contact surfaces.

Fifth, the friction force and the torque consist of several sources. Purohit & Purohit [10] described three of them; the viscous friction torque, the load-dependent friction torque and the seal friction.

Sixth, the bearing damping consists of the lubricant film damping, the material damping (due to the Hertzian deformation of the rolling bodies) and the housing interface damping as described by Sapanen & Mikkola [10]. Experimental measures show that the bearing damping decreases when the rotational speed increases. In the end, the bearing damping effect appears as a bearing damping coefficient in the dynamic analysis equations.

Seventh, the effect of varying of the preload is important. Purohit & Purohit [16] observed that the deflection changes by increasing the preload and also the initial axial displacements change. Thus, with large preloads, the vibration amplitudes associated with the rolling element passage frequency are reduced.

Eighth, the waviness of the bearing rings is one of the causes of bearing induced vibrations such as misalignment and bearing defects. Sapanen & Mikkola [10] described that the modelling of the bearing misalignment is based on the relative displacement and rotations between the bearing rings. Moreover, it can be defined as the initial offset to the relative displacement. However, the waviness modelling requires a specific roundness profile function, for example the Fourier cosine series as suggested in [10]. Thus, the two summations of waviness function (one for the inner ring and the second for the

outer ring) are considered and included in the elastic deformation equation.

Ninth, the slippage in the rolling bearing is defined in terms of a percentage variation of the mean frequency of impact and is usually between 1% and 2% as stated in [13].

Tenth, the defect localization is one of the elementary modelling techniques, in particular, for condition monitoring applications. The defect localisation models are based on the passage frequency of the rolling elements; due to the vibrations associated with the passage frequency of the rolling element. This technique is valid for the spall fault modelling; however it has still some challenges to cope with. Sawalhi & Randall [13] provide a model for the inner/outer race localised fault and the rolling element localised fault based on the depth and angular width of the defect. It is common in condition monitoring research to introduce a specific defect virtually to the simulation model and artificially to the experimental tests. The main distinguishing point between the spall fault and the wear fault is related to the fact that the spall fault has a number of features that are helpful to model compared to the wear fault. First, the spall fault has an approximate dimension which is helpful for the modelling and testing purposes. For example, in [17] it was assumed in their FE model that the width of spall defect is approximately 1.57mm. Also, a defect with 0.8mm width and 0.3mm depth was artificially introduced in the experimental test [13]. Second, the spall fault is localised. Third, it has 'entry departure' fault edges, which has an amplitude effect on the generated signals as described in the double impulse model. Epps & McCallion [18] modelled the double impulse phenomenon that originates from the entry of the roller into the defect area and when the roller departs from the defect area.

3. MODELLING OF WEAR RATE

Wear is generally classified into different mechanisms (i.e. the process by which material is removed from the contact surface). However, since these wear mechanisms are not mutually exclusive, developing a wear prediction based on an individual wear mechanism is not a good assumption for practical applications as has been expressed in [1]. As a consequence of such difficulty to have separate prediction for each wear mechanism, most of the practical wear equations are oriented to the amount of wear (i.e. based on volume or height) rather than based on the individual amount of adhesive or abrasive wear. Even though the literature classifies the most fundamental wear equation (Archard's equation) as adhesive wear equation that is not totally true. It is more accurate to classify it under general sliding wear conditions. The reason is that the wear coefficient is considered as the most important term in the Archard's equation and it is still not clearly traceable to which wear mechanism it is related. The wear coefficient might represent the probability of adhesive, abrasive and other wear mechanisms. Initially, the wear coefficient is considered only as the probability of adhesive wear mechanism. Thus, the wear coefficient is a term of the total probability of all wear mechanisms that might be present under specific contact conditions. The wear coefficient is used to describe the overall wear as an emergent outcome of the different wear mechanisms. Furthermore, it means that under rolling conditions a different wear coefficient should be addressed instead of the ones that are derived by experimental

or theoretical models under sliding conditions. For the rolling bearing, the impact condition should also be considered together with the sliding/rolling contacts conditions. In wider scope, the practical applications show that most of the operational parameters (e.g. applied load, sliding distance, contact geometry, hardness, asperity contact angle, lubrication parameters, etc.) are varying over time. These variations are the main reason related to the scattered nature in the measured wear rates.

There are a number of considerations and delimitations that have been considered in the wear prediction models:

1. The wear equation is the basic element for the numerical wear analysis. Mukras et al. [2] expressed that the fundamental wear equation is the one proposed by Archard in 1953 and various forms have further been developed depending on the intended applications. Archard's equation is a linear wear model that estimates wear based on the applied contact conditions and the operating tribological conditions. The relative sliding distance, the wear coefficient and the contact pressure resulting from the contact are the main parameters of Archard's equation. Generally, the load wear rate diagrams represent the linear relationships and indicate the mild severe transitions as a sharp increase in the slope of the wear rate versus the load curve, please see Zhang & Alpast [19].
2. The relative sliding distance and the incremental sliding distance are the main updated data that are required by the iterative wear procedure. The wear depth and the overall wear depth also require iterative wear modelling procedure as well. In [19] it is described that the volumetric loss of material due to wear increases linearly with the sliding distance indicating that the wear progresses under the steady state conditions. The transition to severe wear was observed at a specific sliding distance.
3. The sliding velocity is one of the complicated issues that have nonlinear effect on the wear rate. Williams [19] described the effect of sliding velocity on the wear rates. The wear rates decrease with the increase of the test speed and reach a minimum value at a certain speed, thereafter, the wear rates start to increase again. This phenomenon is explained so that there is a competition between strain hardening and thermal softening that may occur during the increase of sliding speed. It means that the strain hardening effect is dominant during the first reduction of the wear rate, but later the counterbalanced effect by the softening of the surface layer becomes dominant as a result of the frictional heating.
4. The wear coefficient is the most critical term within Archard's equation due to the assumption of proportionality constant. Often, the wear coefficient is estimated based on experimental studies. However, for the abrasive wear mechanism, Williams [20] proposed an equation where the wear coefficient is related to the average surface slope or the roughness of the abrading surface which carries the array of asperities. Another abrasive wear model is based on the cutting theory where the attack and shear angles of the asperities of the

contacting bodies are the main parameters. Practically, the abrasive wear model based on the cutting theory is in most cases an overestimate of the wear rate due to the assumption that all the displaced material was removed or lost. For oxidative wear in [1], the Quinn's model of oxidative wear is represented. The wear coefficient of the oxidative wear has direct relationship with the sliding distance, the sliding velocity, the fraction of oxygen in the oxide, the oxidation rate, the critical thickness of the oxide layer and the density of the oxide. It is noted that the oxidation rate depends on the activation energy for the oxidation under sliding and the temperature during the contact.

5. The contact force and pressure are estimated according to the procedure that has been explained previously in the dynamic modelling part for restoring contact forces.
6. The transition between mild and severe wear occurred at a critical combination of the applied load, the sliding distance and the sliding velocity as observed by Zhang & Alpast [19].

These consideration issues might be the reason behind Schmitz et al. [21]'s observation "the difficulty to accurately predict component life due to wear is because as reported wear rates are generally exhibit to large range of scatter".

4. ROLLING BEARING CONSIDRATIONS

In general, the measured wear rates (i.e. volume or depth) are scattered as observed by Schmitz et al. [21]. The practical case studies and experimental tests show that some of the previous modelling assumptions are not totally valid for in-site applications over the lifetime of the bearing. Therefore, it is reported by many researchers that wear predictions are still far from reality. In the following, some practical aspects are summarised and discussed:

1. The variations in the applied loading might generate a severe impact at the contact level of the analysis. In [22] the effect of rotational speed fluctuations on the dynamic behaviour of the rolling element bearings were studied. Zhang & Alpast [19] stated that it was commonly observed that increasing the applied speed initially causes a decrease in the wear rate, but later the wear rate starts to increase over time again followed by a transition to severe wear and seizure.
2. The variation of loading conditions such as temperature might directly affect the lubricant and tribology contact and indirectly the lubricant lifetime.
3. The cleanliness of the lubricant is a critical parameter that influences the validity of two-body and three-body models.
4. The variation of the debris interaction rate might have a severe impact on the wear mechanism performance. Its impact highly depends on the amount, shape, sizes, hardness, etc.
5. The surface roughness, waviness and distributed micro-defect might lead to the localization of the wear defect in some specific points within the contact area in which the

assumption of similar wear rate over the whole sliding contact area might not be always valid.

6. Transitions from mild wear to severe wear might occur and represent an exponential increase in the wear amount; however, such transition limits are not addressed in the current wear analysis.
7. Interference vibrations and background noise (i.e. due to machinery induced vibrations) might influence the impact wear mechanism. However, the dynamic models do not address their impacts as a feedback into the wear analysis. It is mentioned that the vibration signal from a single point fault can be expressed as a sum of the combination of the impulses produced due to a specific fault (i.e. including modulation effect, bearing induced vibration), the machinery induced vibrations and the noise which are encountered in any measurement system [17].
8. The lubrication performance is not constant and depends on a number of operating condition parameters and their variation and the degradation over the lubricant lifetime. Moreover, the effect of cavitation is highly important in the elasto-hydrodynamic lubrication. Cavitation is the vaporization of the lubricant or the release of dissolved gas in the lubricant which might cause the starvation of the lubricant or even the breakdown of the lubricating film in the contacting surfaces, please see [23].
9. In the rolling bearing, there are two surfaces required for the three-body contact model at the same time of analysis due to the fact that the rolling element is in contact with the outer ring and the inner ring at the same time. However, only one side of the rolling element contact is considered in the model.
10. In the rolling bearing operations, the occasions of impact loads, sudden stops (i.e. including severe sliding contact) might influence the transition to severe wear.
11. The starvation of the lubricant might occur in frequent occasions. Thus, the starvation events might produce a number of transitions between the tribology regimes of the contact (Hydrodynamic, Elasto-hydrodynamic and boundary lubrication, un-lubricated contact).
12. In a coated surface, there might be some interactions between the coating layer and the substrate layer.
13. In the early stages of wear processes, the wear mechanism might not lead to material removal, it might only produce long-track brinelling and smoothening processes created by frequent rolling element passages. Thus, this stage seems like an incubation stage of the surface wear process.
14. Especially for large scale components, some idling events might occur and introduce static stresses due to the heavy masses and inertia. This leads to produce minor brinelled areas; later these brinelled areas might collect debris and produce severe three-body interactions.
15. In order to obtain accurate results of wear estimation procedures, the procedures should be iterative since wear estimation is highly depending on the effectiveness of the

updated geometry and the accuracy of the contact pressure calculation.

16. One of the phenomenon before material removal by cutting is called ploughing where the material is only displaced and not lost.

In summary, the main assumptions of the current dynamic and wear models are summarized in the following points to trace back the main lack of consideration within the dynamic modelling and wear analysis of rolling bearing.

1. The model utilises a deterministic setting of loading and operating conditions as they apply constant load and constant wear coefficient.
2. The model deals with ideal contact surfaces.
3. The model is based on linear incremental degradation phenomenon as the clearances are predefined and not dynamically updated into the dynamic analysis model.
4. The model lacks to consider the holistic wear degradation process, in the sense that the influence of adhesive, abrasive, and fatigue wear are taken as overall wear and traceable to each mechanism. From the modelling point of view, these different wear mechanisms have different critical parameters, which mean that the model is not able to represent the dynamic influence of these parameters.
5. The model lacks the consideration of the uncertainty estimation for both the modelling and measurements uncertainty.

5. INTEGRATION OF WEAR MODEL INTO DYNAMIC ANALYSIS

The need to integrate the dynamic modelling and wear analysis together is rapidly increasing due to the increase of complexity of modern applications. There are some preliminary interesting efforts that have been presented in [1] for the wear in a high power clutch system and in [2] for the wear in a slider-crank mechanism. The main integration idea is based on the clearance. The assumption is that the wear process reduces the height of the contact surface and that iteratively ends up as updated clearances in the system dynamic analysis. In the same manner, the updated contact pressure and sliding distance that are generated by the system dynamic analysis iteratively are introduced into the wear model.

For the rolling bearing element, most of the research efforts were focused to integrate the system dynamic modelling and the spall fault modelling (i.e. fatigue mechanism) due to its severity on the system health. However, comparing both the spalling and the wear fault cases, two issues are noted:

First, it is clear that the spall models are more oriented towards the detection than the prediction. It is clear that the modelling

and experimental testing procedure introduce an initial spall fault into the simulation or experiment settings as a pre-defined starting point. In other words, this means the spall fault mechanism and its development are not considered in the modelling. From the modelling and the prognostic point of view, excluding the understanding of the fault mechanism within the dynamic modelling might impact the capability to predict and track the fault development process. The detection of a pre-defined severe spall fault might prevent from having severe damage events; however, prognosis in the early fault development stages might provide more cost effective options for the maintenance actions.

Second, some of the industrial applications lose their function due to the wear fault and not directly due to the spall faults. Furthermore, due to the difficulty of detecting the wear fault, the loaded component is worn and influencing the other components within the same assembly and neighbouring assemblies. In this sense, two aspects are important to consider. First, the clearances of the rolling element bearing are extremely important to be as accurate as designed, thus, the clearances of a rolling bearing might increase due to the wear process and lead to performance reduction and influence other components without a real spall fault to be detected. Second, the distributed or micro-pitting faults due to the wear mechanisms are one of the critical seeds to localise a spall fault over running time, in this case, it is always worth to consider wear as an early fault process.

Therefore, the integrated model is conceptually developed as shown in figure 1 and it consists of four modules, as follows:

Module 1: The dynamic operating and loading conditions analysis module that provides loading (load patterns, load fluctuation, duty cycle features, stochastic features, etc.) and operating conditions (temperatures, run/stop ratio, cleanliness, etc.).

Module 2: The dynamic analysis module is based on the dynamic modelling which estimates the applied dynamic forces.

Module 3: The wear analysis basically estimates either wear volume or height for a specific contact area. For example, Archard's wear equation requires, as discussed previously, applied contact pressure and sliding distance parameters.

Module 4: The health analysis module has two parts to deal with the analysis related to the health of the studied system based on e.g. vibration signal analysis, oil/debris analysis, etc. The main input for the vibration signal analysis part is the dynamic force outcomes that are estimated by the dynamic analysis module. In the same manner, the main input for the oil/debris analysis part is the outcome from the wear analysis.

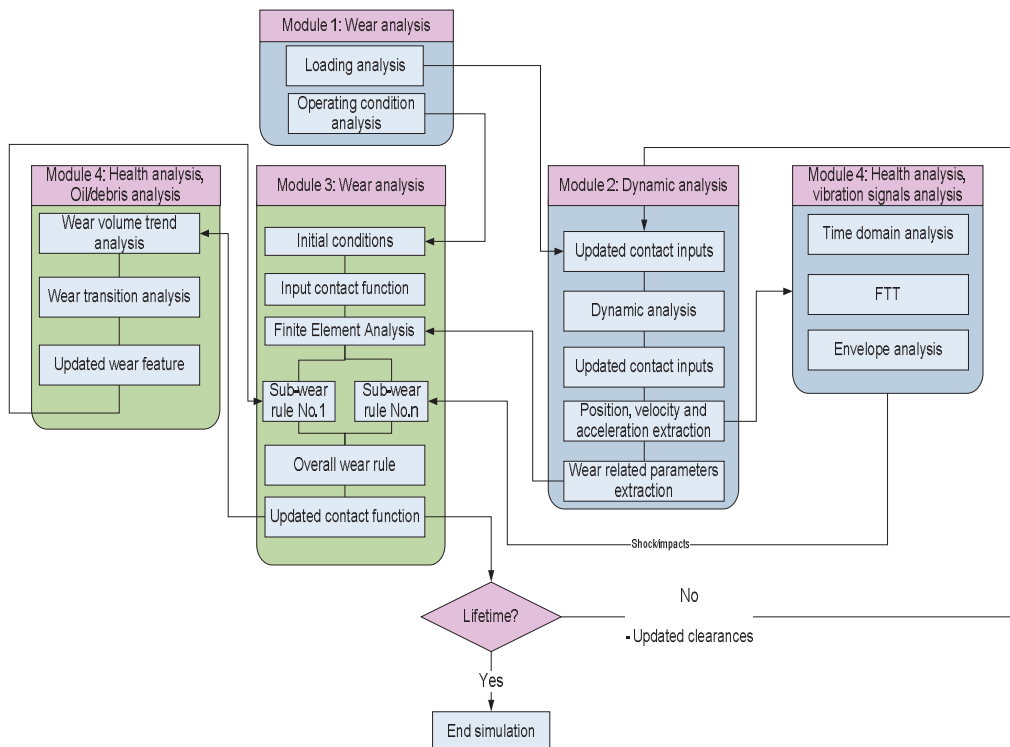


Figure 1. Integrated wear analysis and systems dynamic modelling

6. DISCUSSIONS

The integrated model requires the overcoming of the computational challenges of previous models as listed in the following points:

1. The Finite Element Analysis is a time consuming technique, so the Boundary Element Analysis or simplified techniques such as the Winkler surface model might be faster.
2. The wear predictions require an updated geometry of the contact area which requires numerous cycles of computational analyses as well, therefore, it is popular to use extrapolation techniques.
3. The linking of the system dynamic software (e.g. Simulink) and the finite element software (e.g. ABAQUS) might require a specifically developed interface program.

7. CONCLUSIONS

Wear is a nonlinear phenomenon in terms of occurrence rate and detection ability. That means that the wear rate might have a constant rate for a while, increase exponentially or decrease in a non-linear manner under specific circumstances in a way which makes it difficult either to model or to detect.

It is clear that the most of dynamic and wear models are based on the loading parameters (e.g. applied load, rotational speed) and are

lacking to consider the operating conditions (temperature, poor lubrication, inappropriate lubricant, cleanliness, etc.). Therefore, integrating the dynamic and wear analysis models together on the basis of clearances is not sufficient to produce real wear prediction. Moreover, the rolling and impact deformations should be integrated into the developed wear model.

The integrated model represents a pilot exploration of the modelling requirement of wear process for the rolling element bearing and the conceptual exploration of the potential integration between the systems dynamic and wear analysis model with better consideration of the real loading patterns and operating conditions.

8. ACKNOWLEDGMENTS

The authors would like to acknowledge the VTT graduate school for funding this research.

REFERENCES

- [1] M. Watson, C. Byington, D. Edwards, and S. Amin, "Dynamic Modeling and Wear-Based Remaining Useful Life Prediction of High Power Clutch Systems," *STLE Tribology Transactions*, vol. 48, no. 2, pp. 208-217, 2005.
- [2] S. Mukras, N. H. Kim, W. G. Sawyer, D. B. Jackson, and L. W. Bergquist, "Numerical integration schemes and parallel

- computation for wear prediction using finite element method," *Wear*, vol. 266, pp. 822-831, Mar. 2009.
- [3] A. Palmgren, *Ball and roller bearing engineering*. Philadelphia: S.H.Burbank and Co. Inc., 1947.
- [4] F. J. Harris, *Rolling bearing analysis I*, First. New York: John Wiley & Sons, Inc., 1966.
- [5] P. K. Gupta, "Transient ball motion and skid in ball bearings," *Transactions of the ASME, Journal of lubrication technology*, pp. 261-269, 1975.
- [6] S. Fukata, E. H. Gad, T. Kondou, T. Ayabe, and H. Tamura, "On the vibration of ball bearings," *Bulletin of JSME*, vol. 28, no. 239, pp. 899-904, 1985.
- [7] Y. H. Wijnat, J. A. Wensing, and G. C. Van Nijen, "The influence of lubrication on the dynamic behaviour of ball bearings," *Journal of Sound and Vibration*, vol. 222, no. 4, pp. 579-596, 1999.
- [8] M. Tiwari, K. Gupta, and O. Prakash, "Dynamic response of an unbalanced rotor supported on ball bearings," *Journal of Sound and Vibration*, vol. 238, pp. 757-779, 2000.
- [9] M. Tiwari, K. Gupta, and O. Prakash, "Effect of radial internal clearance of ball bearing on the dynamics of a balanced horizontal rotor," *Journal of Sound and Vibration*, vol. 238, pp. 723-756, 2000.
- [10] J. Sopenan and A. Mikkola, "Dynamic model of a deep-groove ball bearing including localized and distributed defects. Part 1: theory," *Proceedings of the Institution of Mechanical Engineers, Part K: Journal of Multi-body Dynamics*, vol. 217, no. 3, pp. 201-211, Jan. 2003.
- [11] Z. Kiral and H. Karagülle, "Simulation and analysis of vibration signals generated by rolling element bearing with defects," *Tribology International*, vol. 36, no. 9, pp. 667-678, Sep. 2003.
- [12] H. Endo, "A study of gear faults by simulation and the development of differential diagnostic techniques," UNSW, Sydney, 2005.
- [13] N. Sawalhi and R. B. Randall, "Simulating gear and bearing interactions in the presence of faults: Part I. The combined gear bearing dynamic model and the simulation of localised bearing faults," *Mechanical Systems and Signal Processing*, vol. 22, no. 8, pp. 1924-1951, Nov. 2008.
- [14] F. Massi, J. Rocchi, A. Culla, and Y. Berthier, "Coupling system dynamics and contact behaviour: Modelling bearings subjected to environmental induced vibrations and 'false brinelling' degradation," *Mechanical Systems and Signal Processing*, vol. 24, no. 4, pp. 1068-1080, May 2010.
- [15] N. Sawalhi, R. B. Randall, and H. Endo, "The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2616-2633, 2007.
- [16] R. K. Purohit and K. Purohit, "Dyanmic analysis of ball bearings with effect of preload and number of balls," *International journal of applied mechanics and engineering*, vol. 11, no. 1, pp. 77-91, 2006.
- [17] Z. Kiral and H. Karagülle, "Vibration analysis of rolling element bearings with various defects under the action of an unbalanced force," *Mechanical Systems and Signal Processing*, vol. 20, no. 8, pp. 1967-1991, Nov. 2006.
- [18] I. K. Epps and H. McCallion, "An investigation into the characteristics of vibration excited by discrete faults in rolling element bearings," in *Annual conference of the vibration association of New Zealand*, 1994.
- [19] J. Zhang and A. T. Alpast, "Transition between mild and severe wear in aluminium alloys," *Acta mater*, vol. 45, no. 2, pp. 513-528, 1997.
- [20] J. Williams, "Wear modelling: analytical, computational and mapping: a continuum mechanics approach," *Wear*, vol. 225-229, pp. 1-17, Apr. 1999.
- [21] T. L. Schmitz, J. E. Action, D. L. Burris, J. C. Ziegert, and W. G. Sawyer, "Wear-Rate Uncertainty Analysis," *Journal of Tribology*, vol. 126, pp. 802-808, 2004.
- [22] a. N. Lioulios and I. a. Antoniadis, "Effect of rotational speed fluctuations on the dynamic behaviour of rolling element bearings with radial clearances," *International Journal of Mechanical Sciences*, vol. 48, no. 8, pp. 809-829, Aug. 2006.
- [23] T. Otsu, H. Tanaka, N. Izumi, and J. Sugimura, "Effect of Surrounding Gas on Cavitation in EHL," *Tribology Online*, vol. 4, no. 2, pp. 50-54, 2009.

Control charts for assessment of linear asset condition using both temporal and spatial information

Bergquist, Bjarne

Luleå University of Technology
Universitetsvägen 1
SE-971 87 Luleå
+46 920 49 10 00

bjarne.bergquist@ltu.se

Söderholm, Peter

Trafikverket
Box 809
SE-971 25 Luleå
+46 76 792 53 85

peter.soderholm@trafikverket.se

ABSTRACT

To achieve condition-based maintenance, it is necessary to monitor, diagnose and make prognoses about the condition of physical assets. Today, new and innovative technologies for operation and monitoring support the harvesting of large amount of data that may reflect the condition. There are also powerful analytical approaches to refine this data into diagnostic and prognostic information about the current and future condition.

Wear or degradation of a linear asset section is a time dependent process, and a worn or degraded section usually is surrounded by other poor sections. The sampled information will thus be autocorrelated, both in time between consecutive measurements, and along the track in the spatial domain. An effective and efficient decision support procedure for condition-based maintenance actions will thus benefit from taking both temporal and spatial data into consideration.

The purpose of this paper is to present a control chart approach for assessment of the condition of linear assets using both spatial and temporal information. The spatial information stems from the measurements of the asset condition along the track. The track is also split into sections, where section deterioration over time is monitored using information from consecutive measurements in time. The proposed approach is illustrated by a case study of the Swedish Iron ore line, with focus on critical faults of the rail that can cause derailment.

Keywords

Time-series analysis, control chart, condition assessment, linear assets, maintenance, railway infrastructure, Sweden.

1. INTRODUCTION

The railway system is increasingly used for transports of both passengers and freight. Unfortunately, the infrastructure often lacks capacity, which leads to unwanted traffic disturbances. The railway system's capacity can be increased by operating the existing infrastructure in a more efficient and effective way and by developing the maintenance of both infrastructure and rolling stock.

Different actors within the railway system collect large amounts of data. In this paper, it is suggested that this data could be better used for maintenance surveillance and prognosis. Better prognosis and surveillance would reduce maintenance-related stops, and, if the hypothesis is valid, this data carries a large potential to reduce the traffic disturbances caused by infrastructure and rolling stock.

Such an improvement would also reduce the risk level and life cycle cost associated with railway transports.

eMaintenance, that is using innovative Information & Communication Technology (ICT), is key for such improvements, see e.g. [1, 2, 3, 4, 5]. However, the nature of collected condition data usually violate common statistical assumptions, such as samples being taken from an independently, identically and even normally distributed population. The violations of these assumptions differ from being of no consequence to invoking serious errors. It is thus necessary to apply proper analytical approaches that master, for instance data sampled from non-symmetric distributions, and dependencies in the data.

The purpose of this paper is to describe a proposed control chart approach for condition assessment of linear assets, such as transport infrastructure, power lines and pipelines. The approach is illustrated by a case study of twist failure of track at the Swedish Iron ore line.

The remaining part of this paper is outlined as follows. First, there is a short description of the applied study approach, followed by an introduction to statistical monitoring. Then, there is a description of the case study, with a focus on the proposed analytical approach for condition assessment of linear assets. Finally, the paper briefly discusses the results and experiences of the study and suggests directions for further development.

2. STUDY APPROACH

Based on the stated purpose, the following research question was formulated: How can statistical approaches be applied to assess the condition of linear assets? Based on systematic selection criteria [6], a single case study was chosen as an appropriate research strategy to answer the stated research question. Due to accessibility and available resources, it was decided to study the Swedish Iron ore line. The study focused on twist failure of track, since those failures are critical, and had been pinpointed during a criticality assessment.

The empirical data was collected through interviews, document studies, observations and databases. The analysis has been based on theories taken from the quality technology and industrial statistics domains, with a focus on statistical process control. Finally, the paper has been reviewed by key informants and roles in order to verify its content.

3. STATISTICAL MONITORING

Statistically based prediction models are regularly used for predictive purposes. However, the quality of the forecasts is dependent on a number of circumstances. Some examples that have impact on the forecast quality are the quality of the models, the quality of the data fed into the models, and the degree to which the mechanism to be predicted is deterministic, chaotic, or stochastic. Fully stochastic processes cannot be predicted. However, many types of deterioration behaviours are largely deterministic. This means that prediction models can be used to forecast system failures, given that the system conditions and system loads can be measured or estimated.

Statistical process control (SPC) is a classical statistical method used for many surveillance applications to monitor and control that processes run at their full potential. The SPC method was originally developed in the 1920s, see [7], but have since then found use in various sectors [8]. The general SPC methodology is the use of control charts, where measurements of the process is monitored. Using control charts, measurements are compared to nominal or mean values. Any deviation from the nominal or mean is classified either as being small enough that there is reasonably large chance that the deviation is due to random variation (common causes). If the deviation is larger than the control limits, it is deemed too large to be just due to random noise, and thus assignable (due to special causes).

Many types of control charts have been designs for various purposes. For some processes, for instance in manufacturing, it is convenient to sample and measure variables in groups, so called rational subgroups. Automatic measurements, for instance generating measures of all products are, however, increasingly common, and individual observations are also common for variable data. Other control charts include those suitable for attributes, for multivariate data, and so on. The data described in this paper are variables obtained using automatic sampling suitable for individuals charts, and the discussion is from this point limited to control charts for such data.

The selection of control limits is made based on balancing the risks of not detecting an assignable cause, the beta risk. Setting control limits that are too wide will increase the beta risk. A too narrow choice, on the other hand increases the risks of generating false alarms, i.e. the alpha risk. Many regular SPC applications use control limits equal to three standard deviations (3σ), which for normally distributed and independent data that are unaffected by assignable causes for variation would false alarms at 0.3% of the observations. Using individual measurements and using a regular so called \bar{x} -chart with control limits of 3σ would for a control chart of individual measurements, have a chance of detection of 2.3% for assignable causes of size 1σ , and 50% chance of detection of 3σ . See also Montgomery [9].

Observations where assignable causes of variation are present generate alarms so that the process can be adjusted, and so that the assignable causes can be removed. Control charts have been generated for various situations, such as when the data is categorical or numeric, for individual or multivariate properties, for skewly distributed data and so on, see e.g. [9].

Control charts have been used to establish predictive maintenance plans. One example is Katter et al. [10], who use control charts to monitor laser equipment to establish condition-based maintenance of the cathode. Another example is Ben Daya and Rahim [11], who suggest that control charts could be used to monitor

processes where in-control periods are followed by periods with increasing failure rates.

4. CASE STUDY

The Iron ore line is about 500 kilometres long and starts at Narvik in Norway in the north-west and ends at Luleå in Sweden in the south-east (Ofotenbanen is the official name for the Norwegian section). Today, iron ore trains, passenger trains and freight trains use the line. The iron ore transports are performed around the clock throughout the year in extreme arctic climate. Large temperature differences and weather changes are demanding for both the rolling stock and the infrastructure.

The iron ore transports depart from the mining company LKAB's mines in Malmlerget, Svappavaara and Kiruna and arrives at LKAB's iron ore harbours in Narvik and Luleå, as well as a steel plant in Luleå operated by the customer SSAB. The Iron ore line allows a train weight of 8 600 metric tonnes and an axle load of 30 metric tonnes. The annually freights of iron ore are on the northern route (Kiruna–Narvik) 15 million metric tonnes, and 7 million tonnes on the southern route (Luleå–Boden–Gällivare–Kiruna).

The Iron ore line is a bottleneck in LKAB's logistic chain, and hence, the dependability of the line is essential. To minimize transport disruptions, maintenance of vital parts of the railway infrastructure should therefore be preventive and condition-based instead of corrective, to allow timely planning and execution. The condition of the track is fundamental to the railway system, where track failures risk safety and may cause delays due to speed restrictions or derailments. The most critical linear assets of the railway infrastructure are the contact wire and the track. This study focuses on track, due to its criticality with regard to safety and operation, but also cost.

The track condition data used for analysis is collected by a measurement wagon that runs along the track and measures a number of properties in speeds up to 160 km/h. The wagon measures about 30 track geometry variables (every 25th cm), e.g. height position, side position, width, and twist. Measurements of about 40 variables split into wavelength intervals for ripples and waves (every 25th cm), and about 30 rail profile variables (measured every 3rd m) are recorded. Dynamic (with pantograph) and static (without pantograph) measurements of the contact wire position are also obtained (every 2nd and 10th cm respectively). In addition, the ballast profile and track toughness is also measured. See, e.g. [12, 13] for information about the Strix measurement wagon. In this study, the twist is studied, since twist can cause derailment and is thereby vital for safety.

The twist is the difference of cant (in mm) between two sections of track, at distance b (in m) apart, divided by the measurement base b . Hence, twist is expressed in mm/m. The cant is typically expressed as the difference in elevation of the two rails, a quantity referred to as the superelevation. Outside a curve, the two rails should be level, i.e. the cant should be zero. However, on curved track, the cant denotes the raising of the outer rail with respect to the inner rail to allow higher speeds than if the two rails were level. In this case, the cant assists in creating the force necessary to accelerate the train laterally to traverse a curve. At the transition to a curve, the cant increases gradually from zero to the projected cant. If a track was canted to the level required to

generate the full curving force (equilibrium cant) for the maximum speed of the fastest train, a slower train could topple over. A compromise value of cant is therefore used, leading to ‘cant deficiency’ at higher speeds. At the Swedish transport administration’s network, the maximum permissible cant is 150 mm, where the rolling stock is not likely to topple over if they would stop.

If the cant changes too rapidly, there is a risk of derailment. This phenomenon is called twist (i.e. the rate of change of the track superelevation). Normally, the twist is measured on a 6 m base, i.e. the cant measured at two points with 6 m distance. In addition, the twist is also measured on a 3 m base, which usually indicates a low joint in one rail.

The measurement data is stored in a database (Optram) together with information about when and where the measurements were performed. The Optram database also contains information about the infrastructure and its attributes (e.g. type of object, geographical position, description) and if the measurement is taken on a point object or a linear asset. The database also contain information about events and their history, e.g. track alignment with the attributes of machine, method and description, but also the spread with date, track section, and part of track. Optram allows for displaying the measurements and the respective specification limits and for analyses over time, and of parts of the asset. Condition dependency analyses due to factors such as speed can also be studied. The system can also be used for condition predictions of the asset, but does not generate autonomous maintenance suggestions for where different maintenance actions are needed (e.g. track alignment and track grinding). See, e.g. [14, 15] for further information about Optram. The Optram database was used in this study to extract data about the twist and its development along the Iron ore line in both the spatial and the temporal domains.

5. PROPOSED ANALYSIS APPROACH

The measurements of linear assets are often strongly autocorrelated in the spatial domain; an observation of, for instance, track height, will be similar to the track heights measured nearby. This dependency affects the methods of analysis. Without actions to prevent it, the dependency will lead to underestimation of the variation of the process, which in turn will affect prediction properties and evaluations of measurements. Consecutive measurements may also exhibit autocorrelation in the time domain. The measurement of a track height at a certain position will be similar to a measurement taken the next week, or even the next year, given that the track has not been subjected to repair et cetera.

Generally, autocorrelation is handled using two distinct routes when applying control charts [16]. One route is to adjust control limits to compensate for autocorrelation, and the other route is to plot the residuals of a time series model on a standard control chart. In the current paper, a third approach is proposed, which is to sample spatial data over a sufficiently long range. Sufficiently long range does in this case mean that the collected data include all wavelengths of the naturally occurring variation of the measured property. Given that the data is in statistical control and that the sample is sufficiently large, and is including random variation of all occurring wavelengths, this simple approach will render reasonable estimates of the population properties such as

the mean, standard deviation, skewness and so on. In Figure 1, a histogram of spatial observations of the twist failure data is shown, together with descriptive statistics. The 5100 points are sampled from a pool of 510 000 observations; i.e. 127.5 km.

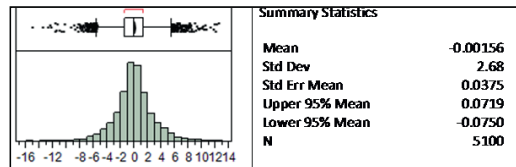


Figure 1. Histogram, stem and leaf diagram, and summary statistics of twist data from track.

Regular control charts for the twist can be designed based on the mean and the standard deviation of the data.

6. RESULTS

A derailment hazardous twist was in October 2010 detected on a section of the Iron ore line track. This alarm prompted a study of the same section using both that measurement data, as well as data from earlier as well as later measurements.

The measurements were analysed in ordinary Shewhart type individuals control charts, based on the sample standard deviation obtained from the analysis seen in Figure 1, see Figure 2 a-f.

Before a control chart is created, it is necessary to select appropriate control limits. A common choice is to use three standard deviations, which, given that the data are normally distributed and independent data would have a false alarm risk of around 1/250. To reduce the possibility for false alarms even further in light of the large data set, wider control limits were chosen. The chosen control limits represent 3.29 times the standard deviation of the data, which would render false alarms of 1/1000. It is clear from the charts in Figure 2 that irregular twisting would have been detected earlier than by the current practice. The charts would signal for an assignable cause more than one year before the measurement requiring immediate actions to adjust the track positions (Figure 2b), and severe deviations four months earlier than the current use of safety alarm limits (Figure 2d). (Range charts are not shown due to size limitations, but the alarms are spotted faster in the individual charts, due to the large spatial autocorrelation).

The repeated measurements make temporal studies of the track section possible. However, a temporal graph requires that the position markers of each measurement are comparable, and in this case the position error was large in relation to the wavelengths of the track twist. To overcome the positioning error so that the twist could be monitored on successive passages of the measurement wagon, the studied track section was split into 50 m intervals. The twist is a property that can be both positive and negative, but if the two rails are to start and end at nearly the same level, the twist must sum to near zero over a longer distance. The range of the twist, thus the maximum minus the minimum twist of the track is therefore used here as a measure of twist problems. The range of the twist variation was measured within each 50m section, assuming that the range of twist within such a section would be a good measure of track twist problems in the section.

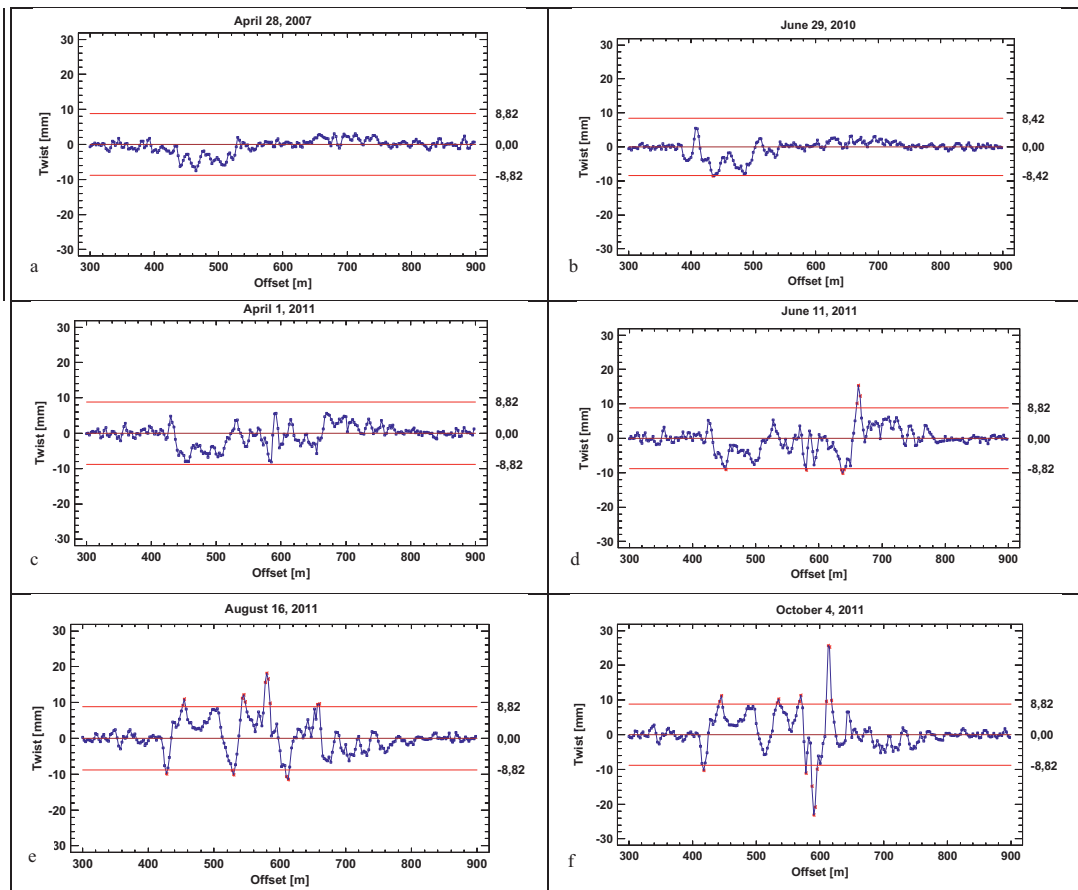


Figure 2. Individuals control chart for twist (6m base) for track section from April 17, 2007 to October 4, 2011. Regular inspection gave alarm of derailment hazardous twist during the latest passage.

Box-Cox transformation test of the twist ranges suggested that the range values should be transformed, and suggested a 95% confidence interval for the power constant to between -0.21 to -0.04. The suggested range was close to a logarithmic transformation (constant equal to zero) and the data was therefore subjected to a logarithmic transformation. A logarithmic transformation is reasonable, since zero is a natural limit for the minimum twist, and the range is skewed to the right. In Figure 3, a histogram and a normal probability plot of the transformed data are shown.

The standard deviation and mean of the logarithmic twist was estimated using data from 17 in-control periods 41625 – 417375, see Figure 3.

In Figure 4, twelve 50m sections are plotted in two charts inspired by the Z-MR charts used for short production runs. A Z-chart lets the analyst plot multiple product types within the same chart, and each product type is plotted in the same chart. Depending on

assumptions, the variation of the different products of a Z chart can be estimated separately for each product, pooled, or scaled. The track sections are indicated through descriptions above the graph and separated by dashed lines. The fifteen observations within each section indicate the fifteen consecutive passages of the measurement wagon during the sampling period. The encircled observation indicates the point where regular procedures detected a derailment hazardous twist fault. Using the proposed method, the section would have rendered an alarm two wagon passages earlier (August 8, 2011). Notice also that the nearby sections also have statistically significant twist, and that these twists seem to be foregone by runs in the temporal domain. Below: Moving range of same data. The observations in the moving range chart are usually well below the alarm limits, due to the temporal autocorrelation present.

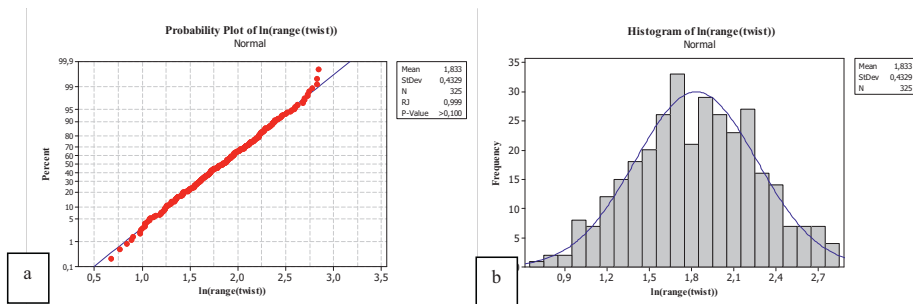


Figure 3. a) Normal probability plot of the twist ranges for 50m sections. b) Histogram of the same data.

An assumption for the observations of this study is that all track sections should be comparable, and thus all sections are plotted versus a common estimate of the mean and average of the logarithm of the twist ranges, see Figure 3. The observations in Figure 4 have then been scaled to unit variance and centred to zero. Each 50m section is delimited by the vertical dashed lines. The data represents 15 passages of the measurement wagon, and each section thus includes 15 observations obtained between April 4, 2007 and June 8, 2012.

The regular procedure with set alarm limits indicated a derailment hazardous twist fault in section 417625 on October 6, 2011. Using the proposed method, the same section would have rendered an

alarm in August 8, 2011. The figure also shows that the track demonstrates twist problems in neighbouring sections. The 417575 section would in this chart be on the verge to indicate problems already in April 1, 2011. Notice also how the twist faults increase over time, which may be used through studying data for runs. Such tests should acknowledge the autocorrelation present, and was beyond the scope of this paper. A final note about the chart: It is apparent that whatever actions that were taken to correct the twist, they did not fully repair the track, as the three consecutive measurements taken after October 6 also had twist problems.

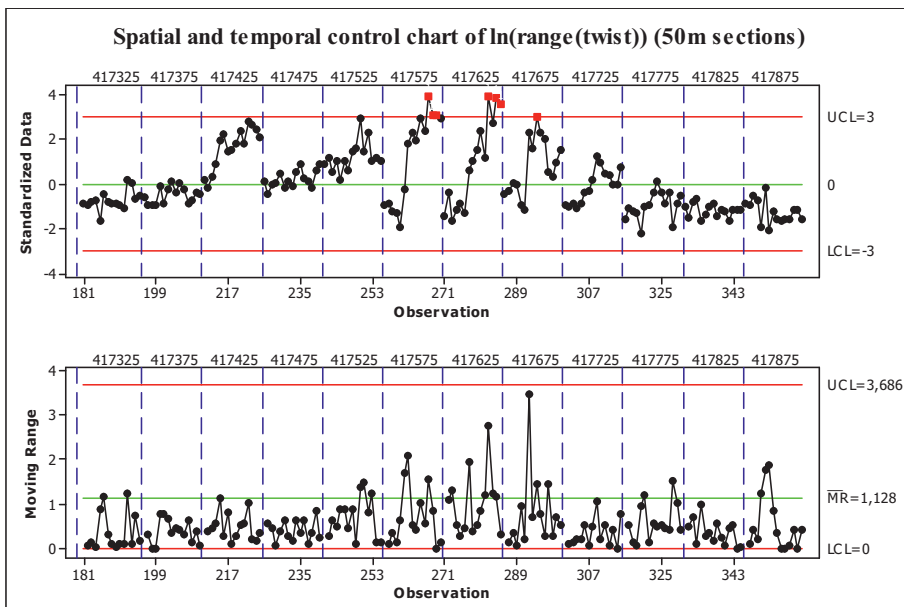


Figure 4. Above: Short run control chart of logarithm of twist ranges per 50m track section.

7. DISCUSSION AND CONCLUSIONS

Two examples of control charts for plotting condition data is presented, i.e. in the spatial and in the temporal domains. The study illustrates how the use of traditional Shewhart control charts in both domains can improve maintenance performance compared to the current practices. This is achieved by the use of statistically based alarm limits instead of relying on maintenance and safety limits that is based on specifications. The study was focusing on one property and on one track section of one line, which is only one type of linear asset, and more studies are naturally needed to test the generalizability of these results.

Admittedly, the procedure to select a large enough sample is ad-hoc and dependent of the judgement of the analyst. A larger error should stem from that the sampling data was based on the whole length of the 127.5 km track, excluding only the section that had rendered an alarm using the regular procedure. The data do therefore contain observations that from a statistical standpoint would be classified as out of control. Exclusion of these points would generate more narrow control limits and even faster alarms, but inevitably also a lower threshold against false alarms. The approach used here is thus to be considered conservative, despite the slightly ad-hoc procedure of estimating the natural variation. Even so, the proposed analysis procedure was successful in pinpointing the fault much earlier than the regular procedure does. However, the analysis procedure including estimation of the standard deviation needs further confirmation to a repeated study, and a study involving more faults and also other properties are needed.

It is indicated here, and intuitive, that statistically based alarm limits are narrower than limits set by mechanistic risks for derailment, such as the current twist limits. More alarms will thus generate more maintenance actions, until the track variation only consists of common cause variation. However, the alarm limits appear to give an early warning, which can be used for planning purposes. Hence, the proposed condition assessment provides a possibility to change the approach from reactive and corrective to a more proactive and preventive one, which allows timely planning of maintenance actions and thereby an avoidance of traffic disturbances.

For this application, we suggest that the regular measurements of the railway condition and geometric alarm limits, should be complemented with statistical alarms. The geometric alarm limits may be useful for de-facto limits of what speeds trains could have on the track, and if the track can be used at all, but are not meant for prognosis of when repair is needed. Statistical monitoring could complement geometrically based alarm limits for the latter purposes, given that statistical alarms have more narrow control limits compared to the geometrical methods.

Even though the proposed control chart approach is illustrated by twist failures of the track, it can be used for any type of track characteristics that are monitored by continuous measurements and result in autocorrelated data, e.g. toughness, geometry and rail profile variables. The condition of other linear assets of the railway infrastructure can also be assessed by the proposed approach, e.g. the contact wire and the ballast. The approach can also be used for condition assessment of other linear assets that may be monitored by continuous measurements, such as road infrastructure, power lines and pipelines.

Besides in other contexts, further studies are needed to thoroughly study the performance of control charts for condition-based

maintenance. The possibility to use more advanced time series analysis for assessment of the condition of linear assets is also an area where research is lacking.

8. ACKNOWLEDGEMENTS

We would like to thank Trafikverket (Swedish Transport Administration) and Luleå Railway Research Centre (JVTC) for supporting this work both intellectually and financially.

9. REFERENCES

- [1] Lee, J. (2003). "E-manufacturing - fundamental, tools, and transformation", *Robotics and Computer-Integrated Manufacturing*, 19(6), pp. 501-507.
- [2] Levrat, E., Iunga, B. and Crespo Marquez, A. (2008). "E-maintenance: review and conceptual framework", *Production Planning & Control*, 19(4), pp. 408-429.
- [3] Karim, R., Candell, O. and Söderholm, P. (2009). "Development of ICT-based maintenance support services", *Journal of Quality in Maintenance Engineering*, 15(2), pp. 127 - 150.
- [4] Karim, R., Kajko-Mattsson, M. Söderholm, P., Candell, O.; Tyrberg, T., Öhlund, H., and Johansson, J. (2008). "Positioning embedded software maintenance within industrial maintenance". In: *IEEE International Conference on Software Maintenance: ICSM 2008*. IEEE, pp. 440 - 443
- [5] Karim, R., Kajko-Mattsson, M. and Söderholm, P. (2008). "Exploiting SOA within eMaintenance". In: *Proceedings of the 2nd international workshop on Systems development in SOA environments: International Conference on Software Engineering. Session: SOA supporting business process management (BPM)*. New York, NY: IEEE, 2008. pp. 75-80.
- [6] Yin, R.K. (2003). *Case Study Research: Design and Methods*, Sage, Thousand Oaks.
- [7] Shewhart, W. (1931). *Economic Control of Quality of Manufactured Product*, Van Nostrand Company, Inc New York.
- [8] MacCarthy, B.L. and Wasuri, T. (2002). "A review of non-standard applications of statistical process control (SPC) charts". *International Journal of Quality and Reliability Management*, 19(3), pp. 295-320.
- [9] Montgomery, D.C. (2008). *Introduction to statistical process control*. John Wiley & Sons, New York.
- [10] Katter Jr., J.G., Tu, J.f., Monacelli, L.E. and Gartner, M. (1998). "Predictive cathode maintenance of an industrial laser using statistical process control", *Journal of Laser Applications*, 10(4), pp. 161-169.
- [11] Ben-Daya, M. and Rahim, M.A. (2000). "Effects of maintenance in the economic design of x-control cart", *European Journal of Operational Research*, 1(1), pp. 131-143.
- [12] Banverket (1997). *BVF 587.02 - Spårålägeskontroll och kvalitetsnormer - Central mätvagn STRIX*. Banverket, Borlänge. (In Swedish)
- [13] Banverket (1997). *Z-647 97-11-15 - Beskrivning av spårålägesystemet i mätvagn STRIX*. Banverket, Borlänge. (In Swedish)

- [14] Bentley (2012). *Bentley Optram*. Available at: <http://www.bentley.com/en-US/Products/Bentley+Optram/>, accessed: 20 August 2012.
- [15] Trafikverket (2012). *Optram*. Available at: <http://www.trafikverket.se/foretag/bygga-och-underhalla/jarnvag/system-verktyg-och-tjanster-for-jarnvagsjobb/optram/>, accessed: 20 August 2012. (In Swedish).
- [16] Stoumbos, Z.G., Reynolds, M.R., Jr. Ryan, T.P. and Woodall, W.H. (2000) "The state of statistical process control as we proceed into the 21st century", *Journal of the American Statistical Association*, 95(451), pp. 992-998.

Information logistics for continuous dependability improvement

Norrbin, Per
SWECO
Box 50120
SE-973 24 Luleå, Sweden
+46 70-6305248
per.norrbin@sweco.se

Söderholm, Peter
Trafikverket
Box 809
SE-971 25 Luleå, Sweden
+46 76 792 53 85
peter.soderholm@trafikverket.se

ABSTRACT

The operational availability performance of technical systems is crucial to provide a quality of services that fulfils the requirements of stakeholders. Hence, continuous dependability improvement is necessary to keep pace with increasing stakeholder requirements and ongoing system degradation to enable stakeholder satisfaction.

Today, information logistics provides methodologies and technologies to support continuous dependability improvement through monitoring of the operational availability performance and its included aspects of reliability performance, maintainability performance and maintenance support performance.

The purpose of this paper is to describe an information logistic framework for continuous dependability improvement that is applied within Sweden to improve the operational availability performance of the railway system. The framework includes an integration of methodologies and technologies from the information technology and dependability management areas that considers both railway infrastructure and rolling stock. The framework is illustrated by a case study of the Swedish iron ore transports, described from the perspectives of the infrastructure manager and the operator of the rolling stock.

Keywords

Information logistics, dependability improvement, availability performance, railway, Sweden.

1. INTRODUCTION

Today, there is an increasing demand to utilise the railway system for transports of both passengers and freight. Unfortunately, there is a lack of infrastructure capacity, which leads to unwanted traffic disturbances. Two ways to increase the utilization of available railway capacity is to make the most of the existing infrastructure in a more efficient and effective way and to develop the maintenance of both infrastructure and rolling stock.

There are also a large amount of data and information collected by different actors within the railway system. If this information could be utilized in a more effective and efficient way for maintenance purposes, there is a potential to reduce the traffic disturbances caused by infrastructure and rolling stock, at the same time as the both the risk level and the Life Cycle Cost (LCC) is decreased.

One key to overcome the challenges and achieve the benefits as described above, is the provision of appropriate information logistic solutions to all stakeholders of the maintenance process through utilization of new and innovative Information & Communication Technology (ICT), i.e. the approach of eMaintenance [1, 2, 3, 4, 5]. However, it is also necessary to apply principles, frameworks, processes, methodologies and tools from the dependability management area. The reason is that dependability management focuses on achieving availability performance of technical systems throughout their lifecycles, which is crucial to provide the right quality of services to stakeholders of the system [6, 7, 8, 9, 10].

The purpose of this paper is to describe a proposed information logistic framework that is applied within Sweden to improve the operational availability performance of the railway system. The framework includes an integration of methodologies and technologies from the information technology and dependability management areas that considers both railway infrastructure and rolling stock.

The framework is illustrated by a case study of the Swedish iron ore transports, exemplified by the perspectives of the infrastructure manager (Trafikverket, Swedish transport administration) and the operator of the rolling stock (LKAB).

The remaining part of this paper is outlined as follows. First there is a short description of the applied study approach. Then, there is a description of the case study, with a focus on the proposed information logistic framework for dependability improvements. Finally, the paper ends with a brief discussion about the results and experiences of the study and gives some interesting directions for further development.

2. STUDY APPROACH

Based on the stated purpose, the following research question was formulated: How can an information logistic framework for dependability management support continuous improvement of the operational availability performance of complex technical systems? Based on systematic selection criteria [11], a single case study was chosen as an appropriate research strategy to answer the stated research question. Due to accessibility and available resources, it was decided to study the current practices at the Swedish iron ore transports, which also enabled the use of action research [12]. Hence, the paper describes some of the context and practices of information logistics and dependability management applied at the iron ore transports. Empirical data has been

collected through action research, interviews, document studies, observations and data bases. The analysis is based on relevant theories and practices, e.g. within information logistics and dependability. Finally, the paper has been reviewed by key informants and roles in order to verify its content.

3. CASE STUDY

The Iron ore line is about 500 kilometres long and starts at Narvik in Norway in the north and ends at Luleå in Sweden in the south. The iron ore transports are performed around the clock all the year in extreme arctic climate, with large temperature differences and weather changes. This harsh operational environment puts very high requirements on both the rolling stock and the infrastructure.

The iron ore transports depart from LKAB's mines in Malmberget, Svappavaara and Kiruna and arrives at LKAB's iron ore harbours in Narvik and Luleå, as well as SSAB's facilities in Luleå. The iron ore line is today the only one in Sweden that allows a train weight of 8 600 metric tonnes and an axle load of 30 metric tonnes. The northern route (Kiruna–Narvik) freights 15 million metric tonnes iron ore per year. The southern route (Luleå–Boden–Gällivare–Kiruna) freights seven million tonnes iron ore per year.

Normally, 10 fully loaded iron ore trains run on the iron ore line per day between Kiruna and Narvik with a total load of 60 000 metric tonnes, which equals a value of 60-70 million Swedish kronor (SEK). It corresponds to about 65 percent of LKAB's total deliveries. From Malmberget to Luleå there are normally five trains per day.

Today, LKAB has 13 of the world's strongest locomotives and is increasing the number to 17 in 2014. The IORE-locomotive is designed for extremely heavy train transports. LKAB's iron ore trains consist of 68 iron ore wagons with a load capacity of 100 metric tonnes each. Each train freights 6 800 metric tonnes of iron ore products and has a length of 750 meters.

The investment in four new double IORE locomotives is strategic in LKAB's development plan and increases the train capacity for iron ore transports to 40 million metric tonnes per year. An increased capacity of the Iron ore line is crucial to achieve LKAB's plan on three new mines and an increased delivery capacity of 35 percent until 2015.

Since the iron ore transports are a bottleneck in LKAB's logistic chain, the dependability of the transports is essential. Hence, to achieve the present and future capacity targets for the transports, the availability performance of vital parts of both the railway infrastructure and the rolling stock has to be managed in a systematic and systemic way.

3.1 Technology, systems and practice for collection and aggregation of operational data

When delays are reported in Sweden, information about train passages are collected from eight different train management systems and compared with the time table. Automatic time reporting is realised in a system that compares the train passages with the time tables. The train passages are coded as E or F events, depending on whether the trains are arriving or departing.

Based on these automatic time reports, information are automatically added about time for a train's arrival, departure or passage at a measurement point, e.g. a station. If there is a timetable connected to the measurement point, the time report also includes automatically added information about the train's deviation from it, expressed as relative time in whole minutes. If the deviation is a delay of more than three minutes, the cause of the delay has to be manually identified and reported as well.

One of the eight train management systems in Sweden is ARGUS, which is located in Boden, along the Iron ore line. The principle of the ARGUS system differs some from the seven other train management systems, since it is based on track circuits, while the other train management systems are based on events in the signalling systems. However, the basic principle for the two different approaches is similar.

In the train management system, flags are placed on the tracks in order to achieve reports about when the trains are moving on the railway network. There are six different flags used, depending on the train's direction of movement and:

- D, train number arrives
- E, train arrives
- F, train departs

Depending on which flag that are used, different events are reported. When a train is planned, it receives a unique train number, which is done and registered in the system Trainplan. When a train is introduced into the train management system, this is done manually by a train dispatcher. All the trains are collected from Trainplan. In order to identify where on the network a train is, ARGUS searches in advance and makes the report with the next signal and measurement point when the train is advancing in the direction of movement.

Arrival reports are created and recorded for every arrival report measurement point that a train passes on a station area. However, it is only the last arrival point that is recorded as the valid one. The reason for this practice is to achieve time reports that are as precise as possible with regard to the actual arrival time as perceived by the passengers. The last time that is recorded is used since the measurement point can be placed on quite a distance from the platform, where the train arrives and passengers are embarking and disembarking. Departure reports are only created and recorded for the first measurement point of departure reporting that a train passes on a station, since this is closest to the actual departure time of the train.

For every station the measurement points that are valid for that particular geographical place is coded into the system, so that the system knows what times to use for comparisons.

For a normal 2, 3 – track station it is the track circuits in the switches that are the measurement points, see Figure 1. Arrivals are registered when the arriving train is leaving the switches' track circuit, i.e. from yellow to red block section. Departures are registered when the departing train is on the switches' track circuit, i.e. from yellow to red blocking section.

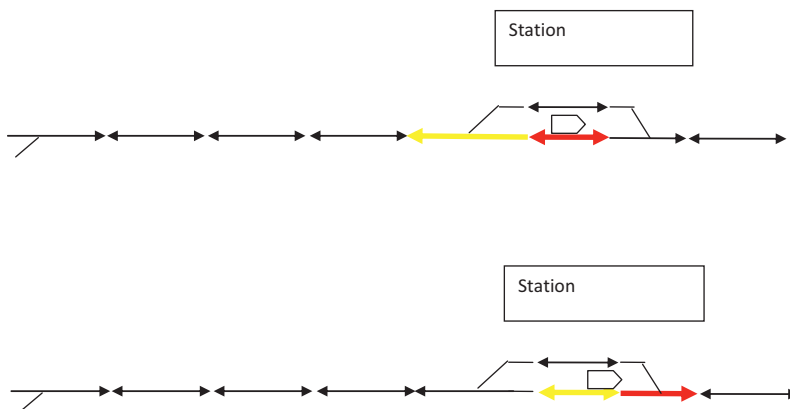


Figure 1. Principle for registration of arriving and departing trains.

Since the report shall contain information about when the train arrives and departs from the platform, the registered times at the measurements points will not coincide with the actual times at the platform. Therefore, the registered times are adjusted before recorded in the ATL system, which correlates times with the trains' location on the railway network. The adjustment is based on the trains' theoretical traveling time between the last measurement point before the platform for arriving trains, and the first measurement point after the platform for departing trains, based on the present speed restriction. Hence, at arrival the recorded arrival time is the time registered at the last measurement

point before the platform with addition of the theoretical remaining traveling time to the platform. At departure, the time that is recorded is based on the registered time at the switch reduced with the theoretical travelling time from the platform to the switch. At larger and more complex stations, there can be some differences compared with the examples given above.

The TPOS system collects all the information from the eight different train management systems in Sweden. In the Trainplan system, the timetables are planned, and once decided upon; the timetables are transferred to the DPP system. See Figure 2.

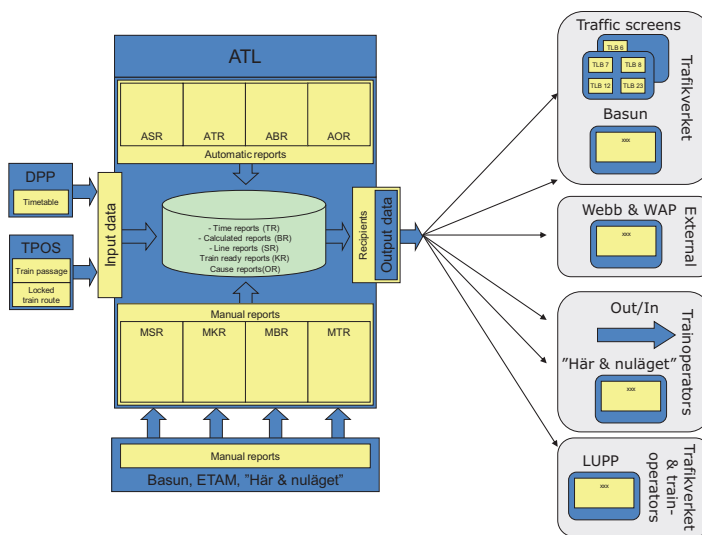


Figure 2. Systems and flow of operational data and information.

The ATL system receives events from the signalling systems from the TPOS system, and timetables from the DPP system and matches these with each other based on configured measurements point in the ATL system, where it is mentioned how the signalling

system events should be interpreted and which adjustment times that should be used in order to compensate for differences in location between used signals and measurement points.

In the two systems Basun and Etam, it is possible to make manual time reporting, which is used when it is not possible to create automatic time reports. The ATL system also receives these manual time reports. Through the Basun system, it is also possible to correct or annul erroneous time reports.

If a time report (automatic or manual) contains a delay of three or more minutes in relation to the successive time report (in timetable order) a message that demands cause reporting is created in the ATL system and is shown in the Basun system, where operative personnel shall make the reporting of causes. These delays and related information is then transferred to the LUPP system for long time storage and analysis.

For the availability performance indicator, delays with cause reporting are used on the highest hierarchical level out of three possible levels. For the availability performance of the infrastructure, causes related to the infrastructure are used. The availability of the rolling stock is calculated in the same way, but uses another set of cause codes.

In the LUPP system, automatic analyses and reports can be created. In the case of availability performance, a number of reports have been created to be used for different purposes. One of these reports summarizes the availability performance of different railway classes of the infrastructure on national level and for the five different geographical maintenance areas, as well as for separate line sections. Another report summarizes the number of cause reported delays for different technical systems of the infrastructure down to separate line sections. These reports can then be used for further analysis or other purposes. One usage is the manually transference to the PULS system, which is used for follow up and analysis of targets, measures and actions related to the balanced score card, see [13] for additional discussion about this usage. Additional analyses may also be made in LUPP or other spread sheet or statistical software, such as Excel or Minitab.

3.2 Operational availability performance measure

Depending on the required quality of service level, a target for the operational availability of the railway system can be selected. This target can in turn be deployed to the railway infrastructure and the rolling stock. Thereafter, the availability targets can be deployed to target levels regarding reliability, maintainability and maintenance support for different required functions of the infrastructure and rolling stock respectively.

When considering the railway infrastructure, these required functions are the track function, the power supply function and the train path function. These three functions are independent of each other, and are all necessary to be present in order to enable the transport service. Due to historical reasons, these functions are in turn addressed to different technical systems of the infrastructure that provide one or more of these functions, e.g. rail, contact wire, and signalling system.

Within Trafikverket, the target levels that are addressed to the functional infrastructure hierarchy are then related to the organizational parts that are responsible for each availability measure. Hence, the organizational parts that are responsible for the technical systems of the infrastructure are also responsible to manage the reliability and the maintainability. Simultaneously, the organizational parts that are responsible to manage infrastructure

maintenance contracts and deal with related entrepreneurs consider the maintenance support.

The availability indicator is on an aggregated level applied for sections, lines and asset classes, where marshalling yards are treated separately, due to their complexity and uniqueness. However, the indicator can also be used on lower indenture levels of the infrastructure, e.g. required functions (i.e. track, power supply, and train path functions) or physical systems (e.g. rail, contact wire, or signalling system).

There are different availability measures that can be applied, see e.g. [6, 7, 8, 9]. In these sources it is seen that availability indicators basically can be calculated as stated in Equations 1 and 2.

$$\bullet \quad A = \text{MUT} / (\text{MUT} + \text{MDT}); 0 \leq A \leq 1 \quad \text{Eq. 1}$$

$$\bullet \quad \bar{A} = 1 - A \quad \text{Eq. 2}$$

where,

- A = Availability (dimensionless)
- \bar{A} = Unavailability (dimensionless)
- MUT = Mean Up Time (measured in e.g. time, distance, or cycles)
- MDT = Mean Down Time (measured in e.g. time, distance, or cycles)

If the time interval of interest is the primary concern, the focus is on instantaneous, limiting, average, or limiting average availability. There are three commonly used availability indicators, i.e. inherent availability (A_i), achieved availability (A_a), and operational availability (A_o). The inherent availability includes the reliability and maintainability of the technical system, which is due to the design of the technical system, while the achieved availability adds preventive maintenance. Finally, the operational availability also adds the support organization's maintenance support performance. See [8] for further discussion about availability measures.

Within Trafikverket, mainly two different types of operational availability (A_o) indicators are applied. One type of indicators is based on clock or calendar time and is used when the MUT and MDT of the required function can be determined independently of the effect on train traffic, e.g. the power supply function and telecommunication. The other type of indicator is based on cycles (i.e. train passages) and is used when MUT and MDT of the required function cannot be easily determined without the effect on train traffic, e.g. the rail function of switches and crossings. The measures for calculating the second kind of operational availability is for MUT the number of successful train passages and for MDT the number of unsuccessful train passages. [13]

A train passage is classified as successful if its delay between two measure points along the considered line is less than three minutes. Hence, if the delay of a train passage is three minutes or more, it is classified as unsuccessful. The time limit for successful train passages can be adjusted according to desire, but the selected level is commonly used within Trafikverket and will also catch the effect of infrastructure failures with small traffic disturbances. Another reason is that train passages with a delay of more than three minutes have to be registered together with the cause of the delay. Hence, the used availability indicator can be deployed from the service level down to the transport system level and its corresponding function, technical system or responsible organizational level.

When aggregating the actual availability during the analysis and follow up, the availability of higher infrastructure levels is based on the geometric mean of the availability of lower infrastructure levels. The lowest level is the distance between two physical measurement points within the infrastructure. The reason for using the geometric mean is that it is more representative than the arithmetic mean when dealing with quotients, which the availability indicator is. In addition to the geometric mean of the proposed availability indicator, a measure of the indicator's spread can be included to not only consider its point estimate. [13]

3.3 Analyses of the operational availability

Normally, the follow-up of the operational availability of the infrastructure is performed on a regular basis every month and more thoroughly every fourth month. However, the follow-up might also be initiated at need, and the interval may be shortened if there are critical risks present. Three common domains for follow up are time, spatial and system. However, even though one domain often is the focus of the analysis, it is normally combined with at least one of the others. [13]

3.3.1 Time domain

Figure 3 shows a trend diagram of the availability for asset classes 1-3 of the railway infrastructure on national level from January 2010 to August 2012. This kind of trend diagram can be used to illustrate the change of the availability performance measure over time on different indenture levels of the rail infrastructure, e.g. geographical maintenance areas, stations, lines and sections.

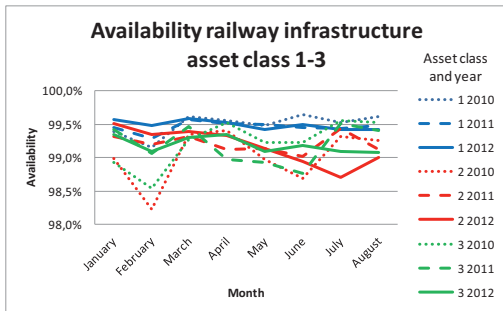


Figure 3. Availability performance for railway infrastructure of railway classes 1-3 from January 2010 to August 2012.
Source: LUPP 10/09/2012.

Figure 3 shows how the availability for asset classes 1-3 changes with time on a monthly basis and between three different years from January to August. As one example, the availability for February is relatively low for both asset classes 2 and 3 during 2010. This winter was quite harsh, which also is seen by a closer examination of the registration of causes of delayed trains, where the failure code of "Snow and ice" is frequent. From a dependability and risk perspective, this can be an indication of insufficient robustness of the railway infrastructure, as well as insufficient performance of the contingency and crisis management of the maintenance organisation.

Table 1 shows another example of how the operational availability may be summarised on both a national level and for different geographical maintenance areas.

	Asset class 1	Asset class 2	Asset class 3	Asset class 4	Asset class 5
Target	99,60%	99,30%	99,30%	99,00%	98,20%
National	99,50%	98,90%	99,20%	98,70%	93,60%
North	N/A	98,40%	99,40%	99,50%	95,70%
Middel	N/A	98,60%	99,00%	98,60%	74,10%
East/Stockholm	99,70%	98,80%	99,20%	98,90%	N/A
West	98,70%	99,20%	98,90%	98,90%	97,70%
South	99,50%	99,30%	99,40%	98,60%	100,00%

Table 1: Summary of the operational availability (%) of railway infrastructure in Sweden for different railway classes on national level and for different maintenance areas during June 2012. Source: LUPP 06/08/2012.

Table 1 shows that the operational availability performance for railway classes 1, 2 and 3 are in line with the target levels (yellow or green colours). However, railway classes 2 and 5 have an availability level that is lower than the target level (red colour).

3.3.2 Spatial domain

Table 1 also shows that the lower availability levels of classes 1 and 2 on national level can be derived to the three maintenance areas North, Middle and East/Stockholm.

When examining the data bases more closely, it is discovered that for maintenance area North (which contains the Iron ore line), all sections of the railway line of class 2 contribute to the low availability level, with the exception of section 116, which has an availability of 99.6%.

3.3.3 System domain

The lowest availability has section 114, which is part of the Iron ore line, with an availability of 96.3%. When addressing the causes of unavailability of this section, measured by unsuccessful train passages to different asset types, it is seen that the top five ones for this section are: overhead wire (147), track (71), positioning system (66), signalling (41) and switches (37). The overhead wire caused damages to multiple pantographs on the line between Luleå and Kiruna, and the wire had to be inspected, during which time trains were cancelled or delayed.

3.3.4 Analysis of the Iron ore line

The analysis of the Iron ore line is divided into the three domains of time, spatial and system (as outlined in the previous section). Figure 4 illustrates the spatial 'time series' plot of the 47 included stations along the Iron ore line and the corresponding boxplot, based on monthly availability performance measures during the time period from January 2010 to July 2012. Hence, each index along the x-axis in the time series plot represents a station along the line, where number 1 is the northernmost station in Sweden (Vassijaure) and number 47 is the southernmost station (Luleå). Each line in the plot represents a specific month during the measured time period. In this plot it is possible to see which stations along the Iron ore line that are bottlenecks from an availability point of view and during which months. For example, station 19 (Harrå) tends to have low availability compared with the other stations, with extremes during October and November 2010 (availability of 80.8% and 80.1% respectively). Station number 5 (Stordalen) also tends to have low availability, e.g. 88.8% during November of 2011. In addition, it is also seen that

Notviken station has a dip in the availability performance during August and September of 2010 (88.4% and 84.5% respectively).

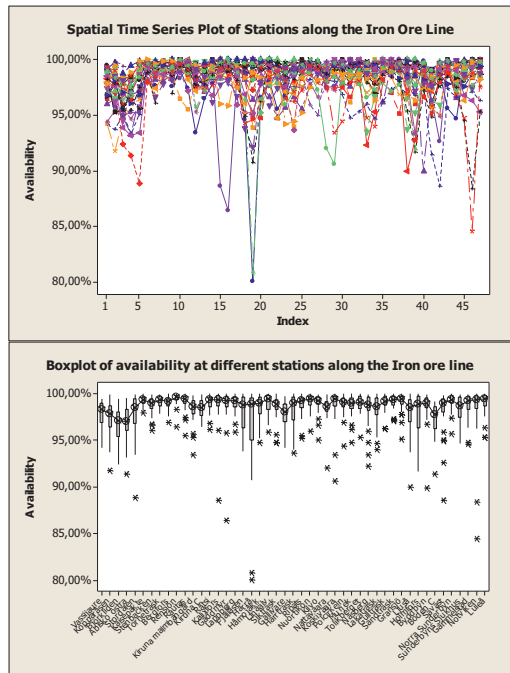


Figure 4. Spatial time series plot of stations along the Iron ore line and corresponding boxplot, based on monthly availability performance measures from January 2010 to July 2012. Source: LUPP 06/08/2012.

Figure 5 illustrates the time series plot of the 47 included stations along the Iron ore line and corresponding boxplot, based on monthly availability performance measures during the time period from January 2010 to July 2012. Hence, the x-axis in the time series plot outlines the time period of the performed measurements as given above. Each line in the plot represents one of the 47 stations along the Iron ore line. In this plot it is possible to see which months that have low availability performance for different stations. For example, it is seen that the lowest availability occurs at October and November 2010 (availability of 80.8% and 80.1% respectively). A closer examination reveals that this is for Harrå stations, as also was concluded by examining Figure 4. In figure 5 it is also seen that there is a dip in the availability performance during September of 2011, which occurs at stations Gäddmyr (86.4%) and Kalixfors (88.6%).

The boxplots (also called box-and-whisker plots) in Figures 4 and 5 can be used to assess and compare sample distributions. In the boxplots it is possible to identify outliers, which are unusually large or small observations, i.e. values that are beyond the so called whiskers. In Figure 4, the x-axis of the boxplot represent different stations along the Iron ore line in Sweden, from Vassujaure in the North to Luleå in the South. In Figure 5, the x-axis of the boxplot represents different dates from January of 2010 (denoted by 1001) to July of 2012 (denoted by 1207). For

example, it is seen in the boxplots of both Figures 4 and 5 that the availability at station 19 (Harrå) during October and November 2010 are outliers. In figure 4, it is also seen that station 19 (Harrå) has larger variation in the availability than the other stations. This conclusion can be drawn by examining the included boxes and whiskers. The boxes include 50% of the data. The bottom of the box represent the first quartile (Q1), i.e. 25% of the data values are less than or equal to this value. The top of the box is the third quartile (Q3), i.e. 75% of the data values are less than or equal to this value. The upper whisker extends to the highest data value within the upper limit, which is calculated as outlined in Equation 3. The lower whisker extends to the lowest value within the lower limit and is calculated as outlined in Equation 4.

- Upper limit = $Q3 + 1.5 (Q3 - Q1)$ Eq. 3
- Lower limit = $Q1 - 1.5 (Q3 - Q1)$ Eq. 4

The boxplots also include a horizontal line, which represents the median of each set of data at the scale markers of the x-axis, i.e. the middle of the data (half of the observations are less than or equal to the median). When considering the median, it is seen in Figure 4 that station 19 (Harrå) does not have the lowest availability performance. Instead, it can be seen that station 4 (Abisko östra) has the lowest availability performance when considering the median of the availability performance.

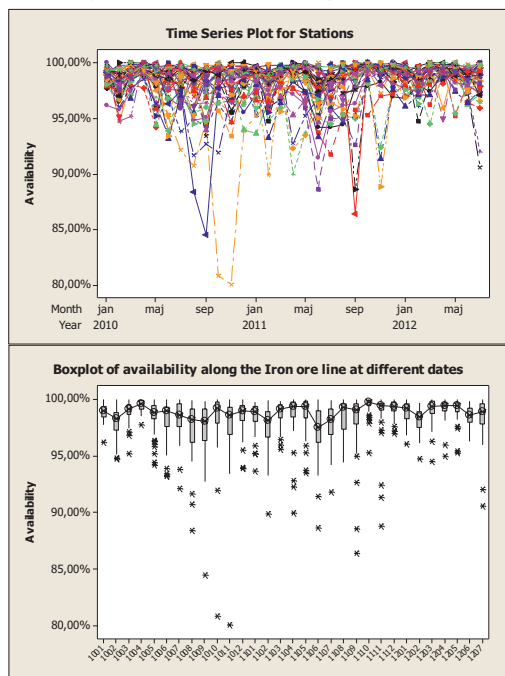


Figure 5. Time series plot of the availability performance of railway infrastructure at different stations along the Iron ore line and corresponding boxplot, from January 2010 to July 2012. Source: LUPP 06/08/2012.

Figures 4 and 5 can give indication where to focus further analysis efforts to solve availability bottlenecks or reduce its variability in

the spatial and time domains. This further analysis can be performed in the system domain.

LKAB is the train operator of the iron ore transports. Hence, LKAB is interested in and analyse the availability of their rolling stock and related train passages together with the infrastructure. This means that LKAB also includes all the causes for unsuccessful train passages, not just the causes related to infrastructure. LKAB's top 10 list of areas with the lowest availability performance in the spatial domain are given in Table 2.

Track section	Station	Availability	Delays	Proportion of delayed trains
116	Svappavaara	74%	5 471	26 of 100
112	Kiruna malmbangård	83%	11 375	17 of 100
114	Koskullskulle	86%	5 121	14 of 100
111	Vassijaure	86%	5 532	14 of 100
122	Luleå	88%	5 172	12 of 100
114	Gällivare	90%	2 481	10 of 100
113	Råtsli	90%	939	10 of 100
111	Björkliden	90%	4 177	10 of 100
112	Kiruna c	91%	1 918	9 of 100
111	Abisko östra	92%	1 349	8 of 100

Table 2: Top ten track sections of the Iron ore line with the lowest availability performance of iron ore transports. Source: LUPP 15/06/2012.

As a train operator that uses Trafikverket's infrastructure assets, LKAB need to have a dialog with Trafikverket on what parts of the infrastructure that are causing unavailability for their transports. A combination of the system and spatial domains can be used to pinpoint the asset types that are causing the most unavailability for LKAB's transports. In Table 3, the top five line sections with regard to availability performance deficiencies are listed, together with the top five assets types for each line section.

Station	Cause	Availability
Björkliden	Switches	98,90%
	Rail	99,70%
	Interlocking system	99,79%
	** (not defined)	99,83%
	Train control centre ARGUS	99,87%
Abisko östra	Rail	99,32%
	Switches	99,41%
	Interlocking system	99,75%
	Positioning	99,83%
	Signals	99,87%
Kiruna malmbangård	Interlocking system	99,66%
	Switches	99,69%
	Train control centre ARGUS	99,73%
	Signals	99,86%
	Overhead wire	99,93%
Vassijaure	Switches	99,24%
	Train control centre ARGUS	99,83%
	Interlocking system	99,83%
	Positioning	99,92%
	** (not defined)	99,92%
Gällivare	Positioning	99,46%
	** (not defined)	99,46%
	Train control centre ARGUS	99,73%
	Switches	99,80%
	Rail	99,86%

Table 3: The top five track sections with the lowest operational availability performance concerning infrastructure together with the corresponding top five asset types causing the unavailability. Source: LUPP 15/06/2012.

4. DISCUSSION AND CONCLUSIONS

The analysis of the availability performance of railway assets may be performed in the time domain, the spatial domain, the system domain, or any combination thereof. The selected domain depends on the purpose of the analysis.

The time domain may be used to identify when dependability deficiencies appear, e.g. seasonal variations or variations during a 24 hour cycle. The geographical domain may be used to identify geographically where dependability improvements are necessary, e.g. specific stations along a line. The system domain may be used to identify which systems or system levels that are crucial from a dependability perspective, e.g. rolling stock, infrastructure, interface between rail and wheel, track, slippers or fasteners. Hence, by combining these three domains it is possible to prioritise dependability improvements by addressing reliability, maintainability and maintenance support issues and adapting available resources to time, place and system.

Independent of domain, the analysis should be made in a deductive (top-down) approach to address deficiencies in the delivered quality of service. The deductive approach starts with examining the availability performance on an aggregated level (e.g. the Iron ore line) and goes step-wise further and further down (e.g. sections, stations and lines between stations) to find root-causes of unavailability related to different types of assets or asset individuals.

In this paper basic point estimates of the availability performance and its distribution are used in a number of different tables and diagrams for analysis purposes. One more sophisticated way to combine estimates of point and spread and relate these to specifications or targets, and present this information in a diagram, is to apply control charts. This is an interesting area for further study. Especially, a multivariate control chart that combines both the spatial and time domains could be a valuable support for systematic dependability improvements.

An interesting visualisation of the control chart would be to use maps of the infrastructure and adding layers representing information from the control chart. For example, different colours and brightness of these colours for different parts of the infrastructure depending on the applied alarm limits of the control chart. An interface based on a map would also enable an easy localisation of bottlenecks or unwanted spread of the availability performance in the spatial domain, and by using different scales it would be possible to zoom in and out to cover different amounts of the considered infrastructure. It would also be interesting to see the development in availability performance over time by being able to display the changes as a movie in a speed that suits the purpose of analysis, e.g. real time for traffic management. A similar application as the last one is already used for road applications in Sweden, when following traffic flows and disturbances in major Swedish cities. Another benefit of a map interface for displaying control chart information tends to be both intuitive and attractive for practitioners.

5. ACKNOWLEDGMENTS

We would like to thank Trafikverket, SWECO and LKAB for supporting this work both intellectually and financially.

6. REFERENCES

- [1] Lee, J. (2003). E-manufacturing-fundamental, tools, and transformation. *Robotics and Computer-Integrated Manufacturing*, 19(6), pp. 501-507.
- [2] Levrat, E., Iunga, B. and Crespo Marquez, A. (2008). E-maintenance: review and conceptual framework. *Production Planning & Control*, 19(4), 2008, pp. 408-429.

- [3] Karim, R., Kajko-Mattsson, M. and Söderholm, P. (2008). "Exploiting SOA within eMaintenance". In *Proceedings of the 2nd International Workshop on Systems Development in SOA environments: International Conference on Software Engineering. Session: SOA supporting business Process Management (BPM)*. New York, NY: IEEE, 2008. pp. 75-80.
- [4] Karim, R., Kajko-Mattsson, M., Söderholm, P., Candell, O., Tyrberg, T., Öhlund, H. and Johansson J. (2008). "Positioning embedded software maintenance within industrial maintenance". In *Proceedings of the IEEE International Conference on Software Maintenance : ICSM 2008*. IEEE, pp. 440 – 443
- [5] Karim, R., Candell, O. and Söderholm, P. (2009). Development of ICT-based maintenance support services. *Journal of Quality in Maintenance Engineering*, 15(2), pp. 127 – 15.
- [6] IEC (1990). *IEC 60050 (191): Dependability and quality of service*. International Electrotechnical Commission, Geneva, Switzerland.
- [7] IEV (2012). *International Electrotechnical Vocabulary (IEV)*. Available at: <http://www.electropedia.org/>, accessed: 18 August 2012.
- [8] Blanchard, B. S. (2004). *Logistics Engineering and Management*. 4th ed., Prentice-Hall, Englewood Cliffs, New Jersey.
- [9] CENELEC (1999). *EN 50126: Railway specifications - The specification and demonstration of Reliability, Availability, Maintainability and Safety (RAMS)*. European Committee for Electrotechnical Standardization, Brussels.
- [10] Söderholm, P. (2007). A system view of the No Fault Found (NFF) phenomenon. *Reliability Engineering & System Safety*. 92(1), pp. 1-14.
- [11] Yin, R.K. (2003). *Case Study Research: Design and Methods*, 3rd ed., Sage, Thousand Oaks, CA.
- [12] E. Gummesson (2000). *Qualitative Methods in Management Research*. Sage, Thousand Oaks.
- [13] Söderholm, P. & Norrbin, P. (2011). A risk-based dependability approach to link maintenance performance measurement and management to overall objectives - A case study within the Swedish railway. In *proceedings of: 1st International Conference on Maintenance Performance Measurement and Management (MPMM 2011)*, pp. 241-247, December 13-15, 2011, Luleå, Sweden.

Holistic Maintenance Information with Multi-layers of Contractors

Mikael Palo Isabelle Lindsund Per-Olof Larsson-Kräik
 Luleå University of Technology Trafikverket
 mikael.palo@ltu.se isabelle.lindsund@hotmail.com per-olof.larsson-kraik@trafikverket.se

ABSTRACT

In Sweden, the railway system is deregulated, with several layers of operators and contractors. By defining maintenance, its objectives and responsibilities, an operation can become cost effective and problem-free. To make this work in a system with many stakeholders, good information flow is crucial. With a holistic computerized information system, each stakeholder can assess maintenance effectiveness, including availability, reliability and maintainability. For this knowledge to be available, the condition of the railway vehicle assets has to be monitored, either manually and/or with wayside detection equipment. If work orders are automatically generated, human factors (i.e., inconsistency) can be reduced. The article uses a vehicle wheel axle as a case study. The maintenance cost for this asset generally represents a large portion of the total maintenance budget; for example, re-tyring a wheel costs about ten times more than re-profiling.

Keywords

Condition monitoring, Holistic information, Maintenance, Railway, Wagon wheels

1. INTRODUCTION

In Sweden, the railway system is deregulated and contains many stakeholders including the infrastructure manager, transport owner, transport operator, train operators, wagon maintenance workshops and wheel maintenance workshop. Each of these layers can comprise several companies; see Figure 1. In addition, each company can have a number of layers. In a system like this, with many layers of contractors and subcontractors, the need for good information flow is crucial.

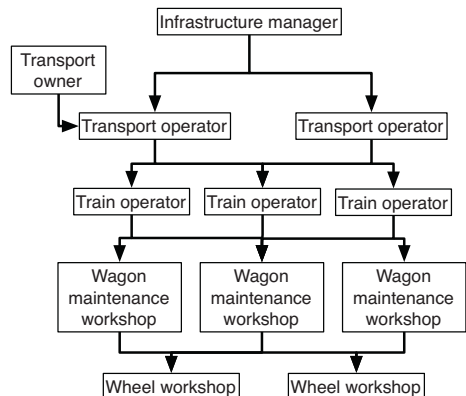


Figure 1. Outline of the deregulated Swedish railway system

One of the most important elements in the dynamics of a railway vehicle is the interaction between the wheel and the rail [7]. The wheel profile determines the stability of a vehicle, and the rate of wheel surface wear [2] determines the life length of a wheel [5]. Increasing emphasis on maintenance

and life cycle costs for both rolling stock and infrastructure have drawn attention to the need to predict wheel and rail wear [9] in order to optimize maintenance decisions and estimations of remaining useful life.

Since the companies are interlinked, determining the total costs of the railway transportation system and creating solutions for its technical problems require the co-operation of all stakeholders. The needs of the railway system as a whole must be put ahead of the needs of the individual companies. The main problem for decision-making in operation and maintenance process is the non-availability of relevant data and information [15]. The recent application of information and communication technology and other emerging technologies facilitate easy and effective collection of data and information.

Given the number of contractors and stakeholders in the Swedish railway, transparent information systems are critical. The contractors have complete responsibility for all aspects of maintenance and maintenance support and must guarantee performance and availability. A clear definition of maintenance, including objectives and responsibilities, is very important for cost effective maintenance and problem-free operation. This is illustrated in Figure 2. Without any formal measures of performance, it is difficult to plan control and improve the outcome of the maintenance process [15]. There is also the problem of how to incorporate Figure 1 into Figure 2 and vice-versa.

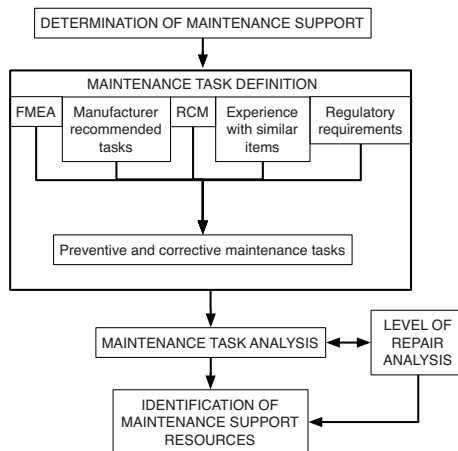


Figure 2. Maintenance and maintenance support planning process

Unplanned failures, such as wheel failures, create disturbances in the system. Disturbances may mean that the planned logistic needs are not met, leading to lost production and delays in delivering products to customers. In the case of large disturbances, the total production may be reduced. As any disturbances in the transport system rapidly disrupt the whole process, it is essential that the maintenance of railway vehicles be planned correctly and performed promptly, effectively and efficiently.

At sharp maintenance peaks, often associated with severe cold and winter conditions, wheel axles become worn more quickly. Since the failures that result from worn axles can have major negative effects, they should be prevented or at least reduced to increase the reliability of the whole transport system.

An earlier study by two of the authors [14] found a large peak of wheels requiring maintenance during winter. Figure 3 shows the increased wheel wear in cold weather; see also [10, 13, 17].

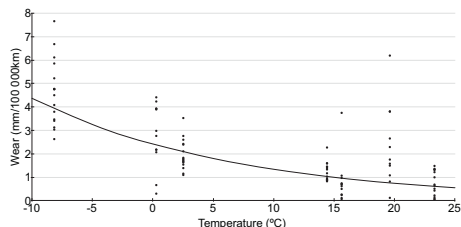


Figure 3. Wear at running circle in mm/100 000 km versus temperature

Optimising wheel axle maintenance must include an economic assessment of acceptable wheel degradation. To this end, the transport owner, along with the various contractors, should consider introducing a holistic computerized information system as this will facilitate their assessment of maintenance effectiveness, including such factors as availability, reliability and maintainability.

The paper is organized as follows. Section 2 will talk about managing conditions and monitoring techniques, as well as wheel deterioration. Section 3 presents the case study; Section 4 provides results and a discussion. The final two sections offer conclusions and suggest directions for future work.

2. CONDITION MONITORING MANAGEMENT

There are different methods to detect and monitor wheel wear and wheel fatigue. One is visual inspection of the wheels at the railway yard. Another is general wagon maintenance overhaul in the workshop. If a wagon with bad wheels is at the workshop, the wheels can be maintained before they have reached their maintenance limit (opportunity based maintenance actions). Wheel maintenance decision criteria are stricter and more rigid at the wagon workshop than at the railway yard.

An important goal of predictive monitoring is to allow early, reliable and cost-effective detection of faults in rolling stock [4], including the early discovery of cracks [8]. An inspection technique traditionally used in the railroad industry is the drive-by inspection, which is not as accurate or reliable as more rigorous and quantitative inspection methods [18]. Condition monitoring uses knowledge of the system of interest to establish its current condition [7]. For example, the railway industry uses wayside detection [2], a technique of detecting specific faults on rolling stock through the use of interrogating sensors placed along the sides of the track, usually at a wayside monitoring station.

2.1. Wayside Detection

The ability to eliminate non-optimally performing equipment by using wayside detectors can directly translate into

enhanced operational safety, improved asset life cycle costs, and increased operating efficiency through fewer unscheduled service disruptions [11]. Wheel condition has historically been managed by identifying and removing wheels from service when they exceed a vertical impact load threshold [3]. These thresholds are typically based on the condition of a wheel/rail impact which is presumed to cause sufficient stress on the track structure. However, wayside detection sites are able to send reports on all passing vehicles, not only those exceeding the safety limits.

Automatic wheel profile monitoring technology uses high speed cameras and lasers to capture the wheel tread profile of each rolling stock wheel as it passes [6]. The equipment monitors wheel profiles against a maintenance standard to detect worn wheels.

Defective wheels which are likely to cause damage to the permanent railway structures can be identified by force measurement detectors; once identified, they can be removed from service immediately [16]. For example, out-of-round wheels can be detected using a wheel impact monitor; these wayside detection systems report impact as either a force at the wheel/rail interface or a relative measure of the defect and are available commercially [2].

2.2. Railway Wheel Deterioration

The interaction between wheel and rail resulting in material deterioration is a complicated process, involving vehicle-track dynamics, contact mechanics, friction wear and lubrication [7]. Two important deterioration mechanisms are wear and rolling contact fatigue (RCF), but even small failures or early cracks are costly, requiring maintenance and possibly causing delays.

Railway wheel flats or slid-flats are a well-known problem in railway engineering [1]. Wheel flats are formed when a wheel set is locked and skids along the rail. The friction between wheel and rail causes the surface of the wheel to become flat instead of round [18].

Spalling is the term used for the RCF phenomenon occurring when surface cracks of thermal origin meet, resulting in part of the wheel coming away from the tread [13]. Cracks from spalling form perpendicular or parallel to the wheel tread surface.

Wheel life can be estimated by examining surface crack initiation and propagation. The propagation rates of surface cracks are non linear with usage, and they increase with the crack length. The depth at which surface cracks appear determines the amount of material that must be removed in the workshop during turning. When a crack reaches a critical size, from a technological, economic or safety perspective, it may be managed and controlled or removed using traditional wheel turning. If wear can be controlled, however, fatigue cracks do not develop and propagate. The philosophy of controlling rolling contact fatigue (RCF) is to manage and control material removal rate.

3. CASE STUDY

The paper studies a transport owner with two different transport operators. Only one operator is part of the data. The operator owns slightly under 1700 wheel axles, but the number is constantly being reduced as wheels become worn and need to be disposed of or re-tired. Nearly 1300 of the wheel axles are used during operation; the rest serve as a backup to prevent stoppages in the system and create transport availability. Overall, the transport system needs an extra 23.5%

axles waiting to be used. The lifetime of a wheel in this system is two to three years depending on operating conditions and maintenance cycle times in the maintenance plant. All wheels are re-profiled at least once a year.

In the case of large disturbances, the total production may be reduced. In addition, the desired availability over time of 90-93% may not be met.

Hence, in the fall of 2011 the company tested a preventative maintenance strategy by adding an extra inspection of the wheels before winter. This was intended to decrease the sharp peak in maintenance during the first cold period of the winter season. A total of 680 wheel axles had a maintenance overhaul. The wheels with wear and wheel fatigue problems were sent for maintenance.

4. RESULTS AND DISCUSSIONS

The number of maintained wheel axles and their maintenance cost for 2011 is split by month and activity. The number of wheel axles maintained during each month also differs by month; see Figure 4.

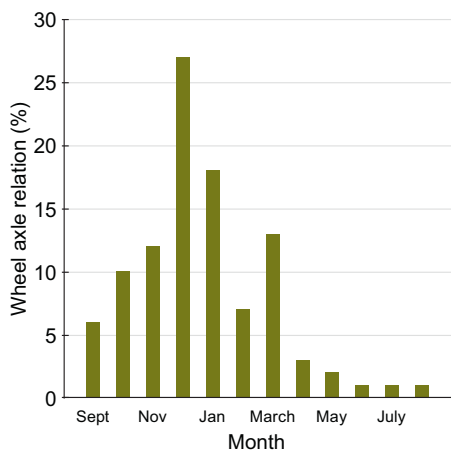


Figure 4. Relationship between maintained wheels per month

Figure 4 shows that December has the most wheel-set failures, while less maintenance is required during the summer months than the rest of the year. This can be due to the winter conditions as well as lower transport volume in summer.

4.1. Maintenance Activity

The data in Figure 4 can be divided into different maintenance activities. In Table 1 these different activities are explained. The relationship between these three activities are presented in Figure 5 and the cost-relation is presented in Figure 6.

Table 1. The different maintenance activities and their explanations

Maintenance activity	Explanation
MA1	Re-profiling
MA2	Re-profiling and bearing revision
MA3	Re-tyre

MA1 is the most significant maintenance activity, followed

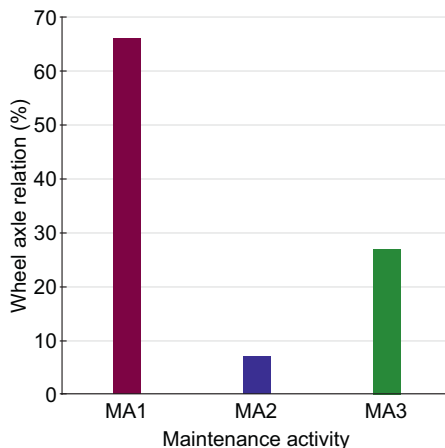


Figure 5. Relationship between the different maintenance activities

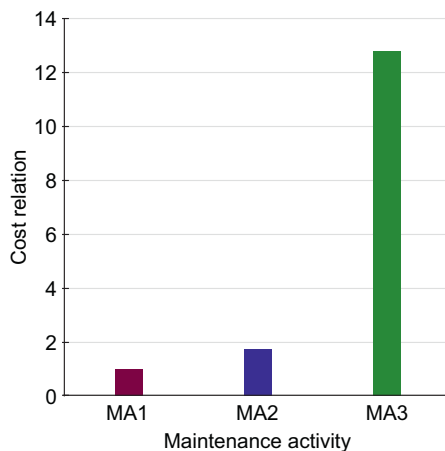


Figure 6. Relationship between maintenance activity and cost

by MA3 and MA2; see Figure 5. The amount of MA2 should be relatively constant over time because it constitutes bearing revisions that are based on travel distance for the wheel axle. The maintenance activities MA1, MA2 and MA3 also differ in cost. The collected data allow us to determine the relative cost of the different maintenance activities, as presented in Figure 6. It is important to decrease the MA3 cost, as it is a cost-driver for the system. A small increase in MA3 will generate a significantly higher cost increase for a similar increase in MA1.

4.2. Failure Modes

There are 35 unique failure modes in the data, all with differing frequencies. The maintenance management system does not contain information on the maintenance activity performed; nor does it link this activity information to the failure modes. Table 2 shows the major distribution of failures.

From Table 2 it is apparent that RCF failures appear in the largest amount, the issues from this is presented in section 2.2. The solution and the significance of it will be discussed

Table 2. The major failure modes for wheel axles

Failure mode	Description	Number of wheels	
331	RCF (whole wheel)	4 788	73%
102	Wheel flat (ice)	479	7%
101	Wheel flat (no ice)	385	6%
591	Brake test	252	4%
330	RCF (Local)	141	2%
Total		6 558	

in the following section.

4.3. Wheel Radius Removal

The difference between wheel profile wear and RCF is that profile wear is a non stochastic process; it is linear and RCF is non-linear with usage. This relation can be seen in Figure 7 [12, pp. 52]. It can also be described as in Figure 8, where it is an intercept; from this point of view, it shows better wear rates than RCF. Interviews with technical field experts indicate that maintenance actions based on profile wear need to remove 5 mm of the wheel radius during each wheel turning, while with RCF, 15 mm or more material must be removed.

Interviews with technical field experts indicate that maintenance actions based on profile wear need to remove five mm. of the wheel radius during each maintenance wheel turning, while with RCF, 15 mm. or more material need to be removed. For the wheels in this study the amount of wheel radius allowed to be removed by turning is about 30 mm, after this the wheel has to be re-tyred.

When turning the wheel for a new profile due to a profile wear, the amount of remaining useful life will be larger than for a turning due to RCF. For a better maintenance economy the aim is to remove the least amount of wheel radius possible. This will provide a longer life length for the wheel and also better overall transport economy.

4.4. Decision Making Process and Evaluation

From the collected data we have identified a *process for decision making and evaluation*. This section contains explanations of which data have been gathered and why they are required; see Figure 9.

Data evaluation The collected data show the failure modes of the wheel and their frequency. This study identifies 35 failure modes. The determination of the failure mode relies on operator-based inspection techniques; these can be subjective, as human factors are involved in the decision making process.

Data on how much of the wheel radius has been removed during each maintenance activity are not available in the existing databases and, hence, are not presented in this paper. This information is desirable for further study; therefore, after each wheel measurement, information should be stored in an available system. Such information could be used to estimate the remaining useful life of the wheel axles; this, in turn, would help in the planning of activities like purchasing or availability.

Missing links The link between failure modes and maintenance actions shown in Figure 9 is not considered in this paper, as this link is hard to extract from the existing databases. Today's databases are separate and independent systems,

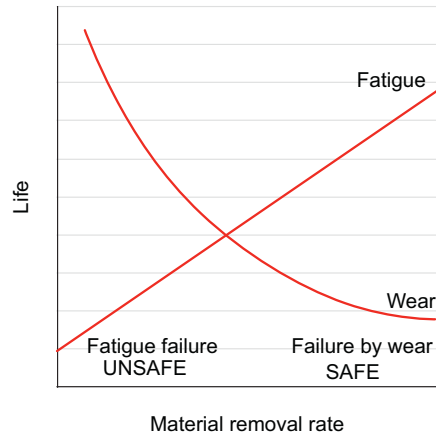


Figure 7. Rail life length versus material removal rate [12, pp. 52]

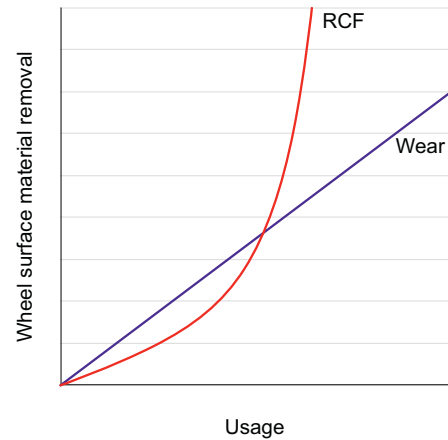


Figure 8. Illustration of the difference between profile wear and RCF

and the interacting companies should cooperate to build a more comprehensive database.

It is not known how much of each maintenance action affects the wheel-set life.

We know how much each maintenance activity costs; by linking a failure mode and maintenance activity, therefore, we can link cost with failure modes. This link will indicate which failure modes are the major cost drivers.

Link and effect decision model Earlier we noted that an increase in MA1 would be beneficial, since the life length of the wheel would increase. Failure mode, maintenance activity, remaining useful life, life length removed by turning, travelled distance and cost must all be derived from the data.

5. CONCLUSION

The data analysis shows that re-tyring costs about ten times more than re-profiling. This indicates that many wheels have a short life length between re-tyring. From an eco-

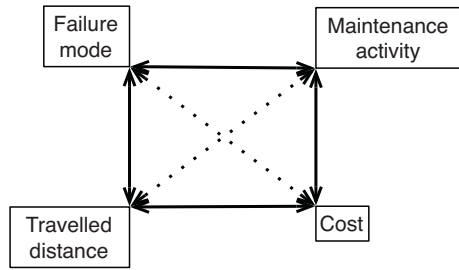


Figure 9. A decision making process and evaluation that can be a basis for analyzes

nomic point of view, this is disastrous in the long run because it leads to large maintenance costs.

The results in this paper show that the uneven yearly distribution needs for wheel maintenance, due to greater damage and wear in winter, means that the wheel maintenance workshop is more heavily used at certain times of the year.

The flow of information can be a problem when there are many contractors. When a number of different companies must exchange data and information, data must be in a common format allowing easy transfer between maintenance systems. If this is not a priority, a great deal of information can be lost; a good solution is the use of a holistic information structure.

A holistic organizational structure will include maintenance data support for decision-making. In this case, high quality maintenance information is necessary to measure and analyse the maintenance performance of the wheel sets and to support the contractors' regulatory requirements. Introducing a holistic computerized information system will facilitate the assessment of maintenance effectiveness, including the availability, reliability and maintainability of assets.

6. FUTURE WORK

The use of condition monitoring techniques and tools should be investigated to evaluate cost and other maintenance limits. Using condition monitoring equipment to decrease the human influence on removing wheel and axles from traffic MA useful when moving into a computerized maintenance management system. Human factors can lead to inconsistency in procedures and quality.

An updated work order system can be used at the sub-contractor level to initiate, control and document specific maintenance tasks according to updated information from the contractors. An updated work request could be triggered automatically by the maintenance information in the system data as the predetermined triggers, such as wayside detector information, calendar time, elapsed time since last task, distance in km, or it could be initiated manually during inspections.

Finally, reported data can include results, observations and resources actually used, thus providing the basis for assessment and improvement.

The maintenance system should be reviewed, including the relationship between the contractors and various economic and technological aspects. The selection of wheels for maintenance and the determination of the failure mode should be standardised for consistency and quality.

ACKNOWLEDGEMENTS

The authors would like to thank Luleå Railway Research Center (JVTC) and partners for their financial support and valuable input.

REFERENCES

- [1] Ahlström, J. and Karlsson, B. [1999], 'Microstructural evaluation and interpretation of the mechanically and thermally affected zone under railway wheel flats', *Wear* **232**, 1–14.
- [2] Barke, D. and Chiu, W. K. [2005], 'Structural health monitoring in the railway industry: A review', *Structural Health Monitoring* **4**.
- [3] Bladon, K. and Hudd, R. [2008], Wayside monitoring of metro lines, in 'The 4th IET International Conference on Railway Condition Monitoring', Derby, UK.
- [4] Bladon, K., Rennison, D., Izbinski, G., Tracy, R. and Bladon, T. [2004], Predictive condition monitoring of railway rolling stock, in 'Conference on Railway Engineering', Darwin, UK.
- [5] Braghin, F., Lewis, R., Dwyer-Joyce, R. and S. Bruni [2006], 'A mathematical model to predict railway wheel profile evolution due to wear', *Wear* **261**, 1253–1264.
- [6] Braren, H., Kennelly, M. and Eide, E. [2009], Wayside detection - component interactions and composite rules, in 'Proceedings of the ASME 2009 Rail Transport Division Fall Conference', Fort Worth, USA.
- [7] Charles, G., Dixon, R. and Goodall, R. [2008], Condition monitoring approaches to estimating wheel-rail profile, in 'Proceedings of the UKACC Control Conference', Manchester.
- [8] Ekberg, A., Kabo, E. and Andersson, H. [2002], 'An engineering model for prediction of rolling contact fatigue of railway wheels', *Fatigue Fract Engng Mater Struct* **25**, 899–909.
- [9] Enblom, R. and Berg, M. [2005], 'Simulation of railway wheel profile development due to wear - influence of disc braking and contact environment', *Wear* **258**, 1005–1063.
- [10] Kalousek, J., Magel, E., Strasser, J., Caldwell, W. N., Kanevsky, G. and Blevins, B. [1996], 'Tribological interrelationship of seasonal fluctuations of freight car wheel wear, contact fatigue shelling and composition brakeshoe consumption', *Wear* **191**, 210–218.
- [11] Keegan, L. S., Blank, R. W., D'Aoust, D., Earl, D. and Ford, D. [2001], Wayside truck performance detector development, in 'ASME International Mechanical Engineering Congress and Exposition', Vol. 21, New York, NY, pp. 157–163.
- [12] Lewis, R. and Olofsson, U., eds [2009], *Wheel-rail interface handbook*, Woodhead publishing.
- [13] Nielsen, J. C. O. and Johansson, A. [2000], 'Out-of-round railway wheels - a literature survey', *Proc. IMechE Part F: J. Rail and Rapid Transit* **214**(2), 79–91.
- [14] Palo, M., Schunnesson, H. and Larsson-Kräik, P.-O. [2011], Maintainability under extreme weather condition, in 'International Heavy Haul Conference', Technical Session, IHHA, Calgary, Canada.
- [15] Parida, A. [2006], 'Maintenance performance measurement system: Application of ict and e-maintenance concepts', *International Journal of Condition Monitoring and Diagnostic Engineering management* **9**(4), 30–34.
- [16] Partington, W. [1993], 'Wheel impact load monitor-

- ing', *Proc. Instn. Civ. Engrs Transp.* **100**, 243–245.
- [17] Sandström, J. and Ekberg, A. [2009], 'Predicting crack growth and risk of rail breaks due to wheel flat impacts in heavy haul operations', *Proc. IMechE Part F: J. Rail and Rapid Transit* **223**, 153–161.
- [18] Stratman, B., Liu, Y. and Mahadevan, S. [2007], 'Structural health monitoring of railroad wheels using wheel impact load detectors', *J. Fail. Anal. and Preven.* **7**, 218–225.

E-Monitoring of Operation and Maintenance Task under Difficult Visual Work Environment

Ravdeep Kour

Division of Operation & Maintenance
Engineering, LTU, Sweden
0046 725521011
ravdeep3123@rediffmail.com

Ramin Karim

Division of Operation & Maintenance
Engineering, LTU, Sweden
0046 920492344
ramin.karim@ltu.se

ABSTRACT

Monitoring objects of interest and obtaining their clear visual appearances are critical requirements for outdoor maintenance tasks and surveillance systems. The difficulties while processing outdoor images are the presence of haze, fog or dust which fades the colours and reduces the contrast of the observed objects. Under bad weather conditions, the light is severely scattered by the atmosphere and images of railway track, coal mine, aviation etc. become degraded. This paper implements a technique for restoring the visibility of weather degraded videos and to detect and track the objects in foggy weather. It has been concluded in the paper that with the application of visibility enhancement algorithm the PSNR for low foggy images is more as compared to high foggy images. Further, the error rate is less for low foggy images. The real time application of the work is to detect objects from the enhanced foggy video that can be used for outdoor operation and maintenance tasks monitoring.

Keywords

Contrast restoration, fog, haze, bad weather, scattering, object detection.

1. INTRODUCTION

Fog has whitening effect on the scenery and drops the atmospheric visibility thus leading to the decline of image contrast and produces fuzzily to the image. The low contrasts of images captured from the video monitor system may sometimes be not so clear to help the railway and the aviation personal to monitor the objects. All these problems can bring difficulties for image information extraction, outdoor image monitoring, target identification and tracking and, therefore, it becomes necessary to enhance foggy image [2, 3]. Object tracking and velocity determination is the process of detecting moving objects of interest, plotting their trajectory by analysing them and then determining their velocity [4]. A monitoring system based on an MSN Messenger that can detect moving objects has been developed. An MSN Auto responder, an authority manager, a detector of moving objects and a streaming media encoder have been integrated into a smart server using the software agent technology to implement the auto-reply and the remote monitoring functions. By using MSN the method do not need a fixed IP for client to real-time reply and do not need extra software or expense [5]. Further, a composite approach for human detection uses skin colour and motion information to first find the candidate foreground objects for human detection, and then uses a more sophisticated technique to classify the objects [8]. Further, in a method to detect and track a

human body in a video, background subtraction has been performed to detect the foreground object, which involves temporal differencing of the consecutive frames and classification of the object is based on two approaches: the first is a codebook approach, and the second involves tracking of the object and if the object can be tracked successfully, it is considered to be a human [7]. Another method uses direct detection of humans from static images as well as video using a classifier trained on human shape and motion features [9]. Further, a method has been used to detect moving objects by computing optic flow only in regions selected by frame differencing [10]. Further, another method uses median filtering to reduce the impulse noise which is the main noise in the coal mine and then the low contrast and illumination was greatly improved with the improved adaptive histogram equalization [12]. This paper analyses the video frames and outputs the location of moving targets within the video frame in real time. The real time aspect addressing conventions, rules and guidelines for the establishment of real-time infrastructure and guidelines for the services requiring real-time execution [11]. The main steps involved in this work are object detection, tracking and analysis of the object. Smart cameras are used as input sensors to record the video. The recorded video may have some noise due to bad weather (fog, haze, dust etc.). Therefore, some pre-processing is required to improve the visibility of the video. Visibility Restoration Algorithm [5] is used for removing noise from the video. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. Frame difference method of segmentation is applied for the separation of foreground and background objects. Feature Extraction plays a major role to detect the moving objects in sequence of frames. Edge detection method of feature extraction is used around the foreground objects produced from frame difference method of segmentation. After the detection of the object, the distance between the initial and final coordinate is obtained. It is stored in the array and the distance is calculated using Euclidean distance formula. The velocity of the object moving from frame to frame is calculated by using the distance and frame rate of the recorded video [1].

2. METHODOLOGY

This work combines two approaches for restoring the visibility of weather degraded images and to detect objects in the videos. The steps followed during the methodology have been demonstrated in the form of a flow chart (Figure 2.1)

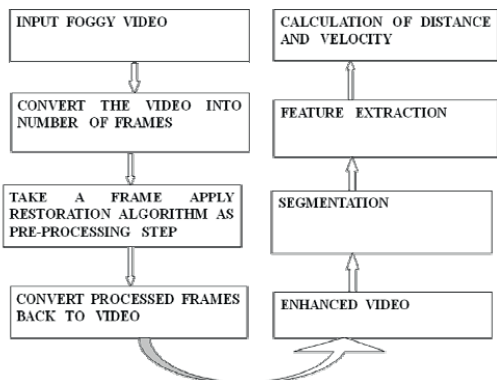


Figure 2.1 flow chart of the steps followed during the methodology.

2.1 Pre-processing

Pre-processing is the first step performed to improve the quality of the captured video using algorithms [5].

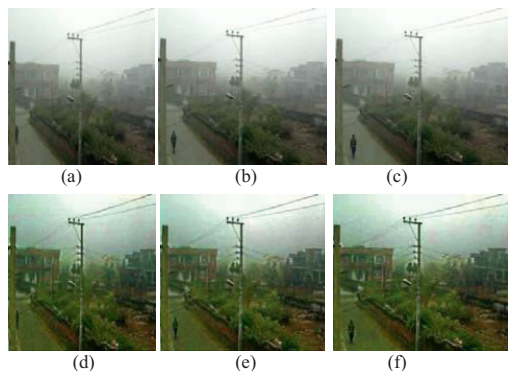


Figure 2.2 Screenshots of video sequence converted to frames, foggy frames (a-c) and enhanced frames (d-f)

2.2 Segmentation

Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. The frame difference method is used for segmentation as it takes less time for execution by calculating the difference between two frames at every pixel position and stores the absolute difference. It is used to visualize the moving objects in a sequence of frames. The following figure 2.3 shows the screenshots of frame difference method of segmentation.

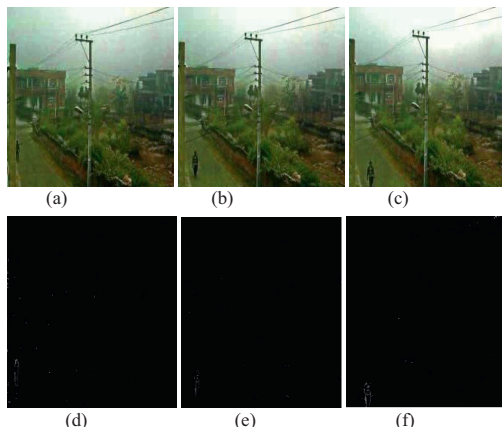


Figure 2.3 Screenshots of the frame difference method of segmentation Enhanced frames (a-d), segmented frames (e-f).

2.3 Feature Extraction

Feature Extraction plays a major role to detect the moving objects in sequence of frames. Extraction of objects using the feature is known as object detection. Every object has a special feature like colour or shape. In this process edge detection method is employed.

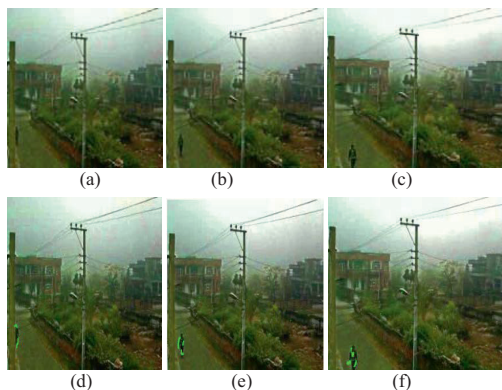


Figure 2.4 The screenshots of the edge detection method of feature extraction restored frames (a-c) extracted frames (d-f).

2.4 Tracking

The process of locating the moving object in sequence of frames is known as tracking. This tracking can be performed by using the feature extraction of objects and detecting the objects in sequence of frames. The distance between the initial and final coordinate is obtained using distance tool in the image processing toolbox and the distance is calculated using Euclidean distance formula. The velocity of the object moving from frame to frame is calculated by using the distance and frame rate of the recorded video and is shown in the figure 2.5.

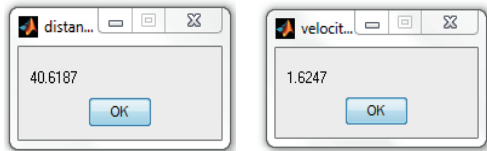


Figure 2.5 screenshots of distance and velocity calculation.

3. RESULTS

Implementations of the algorithm are done using MATLAB image processing toolbox. Object tracking, the major application in security and surveillance is preceded in the following way. In this, a video is recorded using a webcam. The recorded video frames are then converted into individual frames. In the implementation part, fog or noise is removed using algorithm [5]. The filtered images are used as input for the frame difference for the separation of foreground and background objects. Edge detection method is used around the foreground objects produced from frame difference. After the detection of the object, the distance between the initial and final coordinate is obtained using distance tool in the image processing toolbox. It is stored in the array and the distance is calculated using Euclidean distance formula. The velocity of the object moving from frame to frame is calculated by using the distance and frame rate of the recorded video [1]. Figure 3.1 (a-d) shows the Peak Signal to Noise Ratio and Mean Square Error of four different images.



(d) PSNR = 24.139814dB, MSE= 250.668

Figure 3.1 Combination of foggy and enhanced images (a- d) with PSNR and MSE values.

Table 3.1 represents the PSNR and MSE value of Figure 3.1 (a-d).

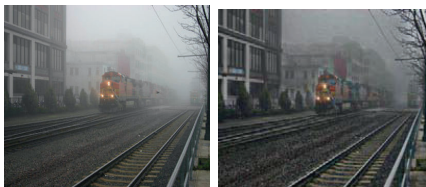
Table 3.1 PSNR and MSE values

Referenced Figure	3.1(a)	3.1(b)	3.1(c)	3.1(d)
PSNR	16.11046	16.7364	20.272	24.13981
MSE	1592.332	1378.596	610.651	250.668

Figure 3.2 reflects the PSNR and MSE value of Fig 3.1 (a-d) which shows the relationship between less foggy and dense foggy images.

Foggy Images

Enhanced Images



(a) PSNR = 16.110467dB, MSE=1592.332



(b) PSNR = 16.7364 dB, MSE=1378.596



(c) PSNR = 20.272872dB, MSE= 610.651

MSE VS PSNR

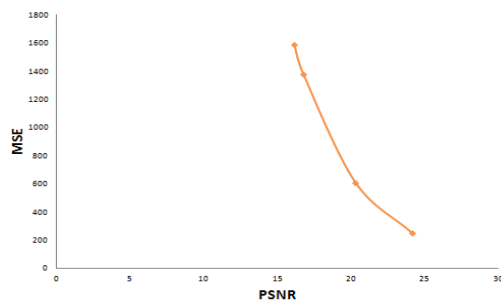


Figure 3.2 MSE VS PSNR

It has been concluded that with the application of visibility enhancement algorithm the PSNR for low foggy images is more as compared to high foggy images. Further, the error rate is less for low foggy images (Figure 3.2).

4. CONCLUSIONS

Object tracking is an important task within the field of computer vision. The proliferation of high-powered computers, the availability of high quality and inexpensive video cameras, and the increasing need for automated video analysis has generated a great deal of interest in object tracking algorithms. Object

tracking from video sequence has various applications; the focus is on tracking moving objects. Object tracking is required in many vision applications such as operation and maintenance, human-computer interfaces, video communication/compression, road traffic control, security, and surveillance system.

5. REFERENCES

- [1] Kour, R., and Dutta, M. (2012). "Detection and Tracking of Moving Objects", *National Conference on Emerging Trends in Computer Engineering and Information Technology* held at National Institute of Technical Teachers Training and Research, Chandigarh, pp. 152-155.
- [2] Kour, R., and Dutta, M. (2011). "Restored Visibility of Images under Fog: A Review", *International Conference on Scientific Paradigm Shift in Information Technology & Management* in collaboration with IEEE, pp. 17-21.
- [3] Kour, R., and Dutta, M. (2010). "Improved Visibility of Images under Fog: A Review", *3rd National Conference on Next Generation Computing and Information Systems (NGCIS)*, pp. 233-236.
- [4] Thombre, D.V. and Nirmal, J. H. (2009). "Human Detection and Tracking using Image Segmentation and Kalman Filter", *International Conference on Intelligent Agent & Multi-Agent Systems*, pp. 1-5.
- [5] Tarel, J. P. and Hautiere, N. (2009). "Fast visibility restoration from a single color or gray level image", *IEEE International Conference on Computer Vision*, pp. 2201-2208.
- [6] Lien, C.H., Chen, P.T., Bai, Y.W., and Ming-Bo, L. (2008). "Monitoring System with Moving Object Detection Based on MSN Messenger", *IEEE Conference on International Instrumentation and Measurement Technology* pp.229-234.
- [7] Zhou, J. and Hoang, J. (2005). "Real time robust human detection and tracking system," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 149 – 149.
- [8] Yoon, S. M. and Kim, H. (2004). "Real-time multiple people detection using skin color, motion and appearance information", *International Workshop on Robot and Human Interactive Communication*, pp. 331–334.
- [9] Viola, P., Jones, M. J., and Snow, D. (2003). "Detecting pedestrians using patterns of motion and appearance", *IEEE International Conference on Computer Vision*, pp. 734–741.
- [10] Elzein, H., Lakshmanan, S., and Watta, P. (2003). "A motion and shapebased pedestrian detection algorithm", *IEEE Intelligent Vehicles Symposium*, pp. 500–504.
- [11] Kajko-Mattsson, M., Karim, R. and Mirijamdotter, A. (2010). "Fundamentals of the eMaintenance Concept", *1st International Workshop and Congress on eMaintenance*, pp. 147-154.
- [12] Zhen-chong, W. and Yan-qin, Z. (2012)"An Image Enhancement Method based on Coal Mine Monitoring System", *Journal of Advanced Materials Research*, Vols. 468 – 471, pp. 204-207.

Assessment of Reliability Data for Traction Frequency Converters using IEEE Std 762 – A Study at Swedish Railway

Yasser A. Mahmood

PhD student

Luleå University of Technology

yasser.ahmed@ltu.se

Ramin Karim

Associate Professor

Luleå University of Technology

ramin.karim@ltu.se

Mustafa Aljumaili

PhD student

Luleå University of Technology

mustafa.aljumaili@ltu.se

Abstract

The IEEE Std 762 is intended to aid the electric power industry in reporting and evaluating electric generating units' reliability, availability and productivity. It was originally developed to overcome difficulties in the interpretation of electric generating unit performance data from various systems and to facilitate comparisons among different systems. This standard was developed to provide terminology and indexes for use in existing data systems, and this data is related to timing, unit states and power and energy capacity. This paper will assess how the data of frequency converters are compatible with the terminology and indexes of the standards. Furthermore, the paper suggests an approach which can be used to generate missing data based on the existing, in order to fulfil the criteria needed to calculate the all-important indexes provided by IEEE Std 762.

Keywords

IEEE Std 762, reliability index, frequency converter.

1. INTRODUCTION

Traction Power Supply System (TPSS) is a provider system to the Electrified Railway by electric traction power. Reliability of TPSS system can be defined as the ability to provide an adequate supply of electrical power without causing safety hazards, time delays or public nuisance [1]. Reliability is one of the most important criteria, which must be taken into consideration during the planning, designing, and operating of these converters. Reliability and availability analysis are always needed to the operational data. In this paper, we will focus on the reliability and availability data of the frequency converter which is considered as an important part of TPSS in the railways. Frequency converter can provide the electric vehicle with adequate traction power. These frequency converters are used to convert the electrical energy from 3-phase, 50 Hz from the public grid to the 1-phase, 16.7Hz traction grid. The converters here in TPSS represent as traction power generators. So, it is acceptable to apply the methodologies of power generator methods like IEEE Std for reliability, availability, and productivity.

The IEEE Std 762 provide a methodology for the interpretation of electric generating unit performance data from various systems and to facilitate comparisons among different systems. It also standardizes terminology and indexes for reporting electric generating unit reliability, availability, and productivity performance measures. The new IEEE Standard "Definitions for Use in reporting Electric Generating Unit (Std. 762)" provides standard terminology and indexes for data systems for performance measurement. A standard is required to formalize methodologies and procedures for unit performance data collection [2-4].

There are some difficulties and challenges to apply these standards to any type of power unit. The main issue of these challenges is related to historical data available for this power unit, as well as the content of these data like the unit state, timing, and power capacity. This paper tries to assess of the converter's data based on the timing, unit state of IEEE Std 762, and how can make the available data of the converter is more compatible of IEEE Std 762.

The rest of the paper is organized as follows: the next section (2) will describe the data of the frequency converter in Swedish TPSS. Section (3) will give an idea about IEEE Std 762 and its important indexes. While section (4) introduce a simple comparison between Ofelia vs. GADS. (5) Describe a program for reading the GELD database. Section (6) will conclude of compatibility between data and Std 762.

2. DATA DESCRIPTIONS OF FREQUENCY CONVERTER

This paper presents a part of collaborative research conducted by Luleå University of Technology for a Swedish Transport Administration (Trafikverket) in Sweden. The operating data was providing from the company in order to study the reliability and the availability of the traction frequency converters. The study is based on operational and maintenance data gathered from the operator of Swedish TPSS within a period from March 2010 to end of March 2012. The data consists of about 130 converters distributed with 46 converter stations around the whole traction network. In order to monitor and statistical recorded for operation, maintenance, and control. There are two kinds of database are available in the Swedish Railways Administration (Trafikverket) for traction frequency converters: Ofelia and GELD.

"Ofelia" is a fault reporting database that gathered data include all significant events for the systems, i.e. date of events, description of fault, the failed component, the description of maintenance, and the repair time, maintenance release times, corrective maintenance, etc.... Generally; it covers all mechanical and electrical TPSS in the Swedish railway. When a fault appears in the Swedish railway infrastructure, it is indicated in the Centralized Train Traffic Control (CTTC) office by alarms from the signal and detector systems. Afterwards, the CTTC office staff initially reports the occurrence of the fault in Ofelia, after that they call the contractor's contact person, who in turn instructs personnel to remedy the fault. Maintenance personnel receive a work order to check the alarm and perform repair work, if necessary, and then drive to the fault location. After the visit on site and the repair, the maintenance workers must report back to the CTTC office, then finalizes the report in Ofelia [5].

"GELD" database contains the readings of electrical measurements like input/output converter current, input/output

station current, station active & reactive power and station input/output power. All of these readings are coming to the control center and recording in text files for all instruments daily. GELD database contains around a tenth of thousands text file for each converter per year, and there is a one file for each converter per day, and this file contains a readings in each 10 minutes.

3. DATA QUALITY ISSUES IN ONVERTER'S DATA

Data quality (DQ) issues in general can be related into the source of the data. The source of the data could be subjective (human) or objective (machine) where the major issues are caused by subjective sources of data [6]. Strong et al [7] have observed the key DQ problems. Their observations were based on research that employed qualitative data collection and analysis techniques. Some of these problems are: 1) Multiple sources of the same information produce different values. 2) Information is produced using subjective judgments, leading to bias. 3) Large volumes of stored information make it difficult to access information in a reasonable time [7].

During our study of converter's data, the major challenges on data collection are related to incomplete and conflicting timing. As an example sometime the recorded repair times do not represent the real repair time. In addition, some data are ambiguous, and non-meaningful. Moreover, there are minor and major faults recorded in the database, and sometimes the minor faults not lead to a service outage of converter. This lack of information can be considered as one of the main issues of this database. As well as there is no mentioned about the current state of the unit and its timing (the time spent in each state), that means; there is no information about the utilization factor or operating hours for each converter. Without operating hours for each converter, we couldn't also use the MCF as reliability and availability indicators.

4. IEEE STD 762 FOR RELIABILITY, AVAILABILITY, AND PRODUCTIVITY

The IEEE standard was developed in 1987, based on efforts started in 1980, to provide terminology and indexes for use in existing data systems. Measures of generating unit performance have been defined, recorded, and utilized by the electric power industry for over 60 years. The increased focus on generating unit performance in a competitive marketplace has made the regulatory agencies and the industry to place a greater emphasis on performance measures. The standard provides a methodology for the interpretation of electric generating unit performance data from various systems and facilitates comparisons among different systems. It also standardizes terminology and indexes for reporting electric generating unit reliability, availability, and productivity performance measures.

There are some definitions according to Std 762: *Reliability* in this standard encompasses measures of the ability of generating units to perform their intended function. *Availability* measures are concerned with the fraction of time in which a unit is capable of providing service and accounts for outage frequency and duration. *Productivity* measures are concerned with the total power produced by a plant with respect to its potential power production. A plant could comprise a unit or a number of units. Therefore, productivity measures consider magnitude of an event as well as frequency and duration of the event.

Any event preventing the ability of the generating unit to produce electricity at its maximum capacity is covered in the scope of this

standard. As well as the standard was developed for application at the unit level; it does not address applications at the plant component or system level.

5. The unit state

To perform the main functions of Std 762 for measuring and evaluation in reliability, availability, and productivity, it needs a sufficient data with categories (unit states, timing, and capacity levels). So, the suitable indexes for all types of units should be calculated: A unit state is a particular unit condition that is important for collecting data on performance. There are two primary unit states: available or unavailable, which defines these primary unit states. These two states are mutually exclusive and exhaustive. A unit should be in exactly in one of these states at all times. Thus, these states divide calendar time into no overlapping segments. The available and unavailable states are divided into additional, mutually exclusive states such as in service, reserve shutdown, planned outage, and unplanned outage. There are many states to define the current state of the power unit in order to apply Std 762 and to calculate the required reliability indexes (see figure 1). Time and dates represent the spent time in the various unit states, some indexes are based on period hours (see figure 1). Capacity terms: Terms that involve capacity can be expressed as gross or net quantities such as Maximum Capacity, Dependable Capacity, Available Capacity, Unit De-rating, Basic Planned De-rating, and Unplanned De-rating.

In 1981, Fernihough suggested a code for all of IEEE states in the database of the power unit in order to make the online monitoring of the current state. These code systems are used in the reporting form or for interactive computer data entry procedures [8]. The North American Electric Reliability Council (NERC) has been operating the Generating Availability Data System (GADS) since 1982. GADS uses IEEE Std 762 definitions for Use in Reporting Electric Generating Unit Reliability, Availability, and Productivity as its primary definitions document. Table 1, shows a sample of raw data of GADS [3]. Table 1 also shows the important two issues regarding starting reliability. The first one is the number of actual starts and attempted starts. The second issue regarding to running reliability like Service Hours (SH), Revers Shutdown Hours (RSH), Availability Hours (AH), Forced Outage Rate (FOR), Forced Outage Events (FO Events). All of these information's should be available online for each power unit in the station.

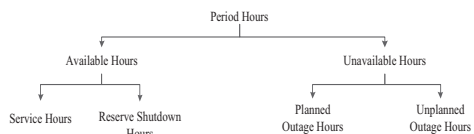


Fig 1: Time spent in various unit states.

There are only two indexes that can use it at two states; Availability Factor (AF) and Unavailability Factor (UF) [9], because we have two states only: available state and unavailable state, according to the 0Felia. When there is a failure, the converter is unavailable, otherwise it's available. Figure 1, shows the time spent among various unit states for the converters, there are some donations about the following states:

- A- Available: The available state is where a converter is capable of converting service, regardless of whether it is actually in service and regardless of the capacity level that can be provided.
 1. In service: The state in which a converter is providing output power to the TPS system.
 2. Reserve Shutdown: The reserve shutdown state is where a unit is available, but not in service.
- B- Unavailable: The state in which a converter is not capable of providing power because of a forced outage or planned maintenance.
 1. Planned Outage: The state in which a converter is not providing output power because of planned maintenance performed on a subsystem or component during a time period, it is needed for the system to produce output power.
 2. Forced Outage: The state in which a converter is unavailable due to catastrophic component failure resulting in loss of or inadequate output power.

Table 1: Raw data used in NERC GADS

Unit	Capacity (MW)	SH	RSH	AH	Actual Starts	Attempted Starts	FOH	FO Event
41	100	183	8576	8759	35	35	1	1
42	150	198	8562	8760	31	31	0	0
43	125	186	6867	7052	37	38	9	2
44	170	105	4128	4233	29	29	4528	3
45	180	62	8259	8319	20	20	98	1
total	725	734	36392	37123	152	153	4636	7

We should note here, we are going to talk more about the Std 762, the indexes that related to the unit alone, that means that we will not mention the weighted, equivalent, demand, and de-rating indexes. There are common indexes like Availability Factor (AF), Service Factor (SF), Forced Outage Factor (FOF), Forced Outage Rate (FOR), and Generation Capacity Factor (GCF) [10].

6. COMPARISON BETWEEN OFELIA AND GADS

The objectives of Std 762 relate to the gathering and presenting of broad system operational data on a consistent basis. It has many conditions related to the timing, transition states, and power & energy capacity to perform the calculations of reliability indexes. See the above two sections, Ofelia database doesn't meet with these conditions. Ofelia database is just a failure reporting, and the reporting process completely depends on the two states, available or not available. It's difficult to know from this database if there is no fault in the unit, is the unit in service state or reserve shutdown state? This is the great deference between a failure reporting systems like Ofelia and what Std 762 data need like GADS, which show the state and its hours for each unit. We can say that there are problems to apply the information which we have in Std 762, and the problem lies in the state base.

In addition to the problems of unit state, not all the recorded failures in Ofelia lead to outage of converters because there are minor and major fault, and the minor faults doesn't lead to outage, and this is not completely mentioned in Ofelia database due to incomplete problem. The most important things in available hours are the service hours and the reserve shutdown hours (see figure 3), and this information also is completely not available in Ofelia. GELD database contains a lot of information related to amount power converted from each converter and current, and the values of power and currents within each 10 minutes are recorded for

each converter. If there is a production current from the converter that means the converter now in service state, and when the current and the output power from this converter is zero, that means the converter is in a reserve shutdown state. To solve this problem and get information from GELD to specify the service state and reserve shutdown state from available state, we designed a program by using m-file Matlab package to read tenth of thousands of files for each converter from 130 converters work in TPSS (GELD reader).

7. GELD PROGRAM READER

The challenge now is how to take the benefit of these two databases, and solve all of difficulties that related to the operation data and the unit state of converters. To solve this problem and get information from GELD to specify the service state and reserve shutdown state from available state, we designed a program by using m-file Matlab package to read tenth of thousands of files for each converter from 130 converters work in TPSS. Then it gives the operating hours (service hours), OFF hours (reserve shutdown hours), and the converted power before fault.

The flow chart of the program is shown in figure 2. It needs initially to select the station name of frequency converter which we need to read its data files, and then select the unit number because there are maybe two, three, or four converters in the station. It also needs to select the year of data which we want to read. There is a form of the name of text files, and this name depends on the short of station name, unit number, type of parameter (power or current), and the date (year-month-day). This is a sample name for the current of converter 2 from Boden station like "2011_01_01_BDN_OMR_2_STRÖM", the abbreviation of OMR refer to the static converter while OMF to the rotary converter. There are a huge number of text file for each month folder, as example there are around 52,888 text files in a folder of January 2011. The month folder contains all the electrical readings for all 130 converters.

When we know the service hours within available time hours, we can apply many reliability indexes like Service Factor (SF), Forced Outage Factor (FOF), and Force Outage Rate (FOR), and also can use the MCF after knowing the operating hours. Hence, the unit states for Frequency Converter will be three states: service, reserve shutdown, and unavailable states (see figure 2).

In addition to the proposed software, there are other ways to solve the difficulties of data; interviews and documents. The interviews were performed with experienced practitioners at Swedish Transport Administration (Trafikverket) including both project and field technical persons in Operation and Maintenance departments. The interviews and discussions supported the research to solve the challenges faced on the data collection, filtering and validity. The documentation consisted of different descriptions, policies, and procedures pertaining to operation, maintenance and reliability analysis of the TPS system and converters, as well as documents and standards supporting reliability analysis of TPS such as IEEE and IEC, etc.

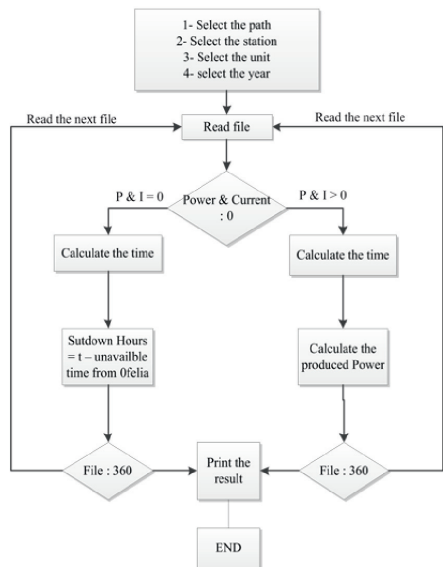


Fig 2: Flowchart of the GELD reader program

8. RESULTS

The main function of the program is to make a survey for all the recorded electrical measurements and calculate the service hour, reserve shutdown hours, and converted power before each failure in each converter. The failure point is not recorded in GELD, also the not working case is not recorded to know this not working due to failure or reserve shutdown. Therefore, for these three issues we should take these information's from Ofelia and compared with the findings of GELD reader. Table 2, shows the reliability, availability, and productivity indexes for 6 converters as a sample of the results. SF and AF for availability while FOF and FOR for reliability and Capacity Factor (CF) for productivity index.

Table 2: Sample of the results of frequency converters

Unit	SF	AF	FOF	FOR	CF
Emmaboda 3	0.2122	0.9932	0.0067	0.0309	0.2816
Kil 3	0.1852	0.9268	0.0731	0.2831	0.3094
Kil 4	0.821	0.9934	0.0065	0.00785	0.5851
Mora 1	0.7527	0.9995	0.0004	0.0005	0.1699
Mora 2	0.7303	0.9997	0.0002	0.0003	0.1552
Ostersund 3	0.0882	0.9729	0.027	0.2346	0.2513

9. CONCLUSIONS

There are two available data for the traction frequency converters, and these data are not meet with what IEEE std 762 need from database to measure the reliability, availability, and productivity. the big differences are related to the unit state and the timing of these unit states. Only two unit state is available in Ofelia database: available and unavailable, there are no mentioned about service time and reserve shutdown time in Ofelia. This type of database is not useful for reliability and availability analysis. There are only

two indexes can use for these information: Available and Unavailable Factors. As well as there are some problems in Ofelia reporting database; not all failures are leading to outages, ambiguous, and non-meaningful.

GELD reader has been built as a program by using Matlab software package to read the GELD database. This database is a recorded electrical measurements, currents and power for the converters, to find the service hour, reserve shutdown hours, and amount converted power before fault. With comparing the results with the failure time in Ofelia failure database to get the spent time at each state. The GADS is a database used with NERC as example of how the data reporting system of power units is used to measure the reliability and availability performance.

10. ACKNOWLEDGEMENT

The authors would like to thank the Swedish Transport Administration (Trafikverket) for providing the data used in the analyses described in this paper. In addition, the authors would like to express their appreciation to Mr Anders Bülund and Mr. Niklas Fransson for all the support and assistance which they have provided during this study.

11. REFERENCES

- [1] S. Sagareli. Traction power systems reliability concepts. American Society of Mechanical Engineers, Rail Transportation Division (Publication) RTD 27pp. 35-40. 2004.
- [2] Anonymous IEEE standard definitions for use in reporting electric generating unit reliability, availability, and productivity. IEEE Std 762-2006 (Revision of IEEE Std 762-1987) pp. C1-66. 2007.
- [3] North American Electric Reliability Corporation (NERC), "Gads data reporting instruction," NERC, http://www.nerc.com/files/Appendix_F_Performance_Indexes_and_Equations.pdf, 2012.
- [4] G. M. Curley. Power plant performance indices in new market environment: IEEE standard 762 working group activities and GADS database. Presented at Power Engineering Society General Meeting, 2006. IEEE. 2006, .
- [5] Stefan Niska, "Measurements and Analysis of Electromagnetic Interferences in the Swedish Railway Systems," Luleå University of Technology, 2008.
- [6] M. Al-Jumaili, V. Rauhala, P. Tretten, R. Karim" Data Quality in eMaintenance: A Call for Research" MPM, 2011.
- [7] Diane M. Strong Yang W. Lee Richard Y. Wang "10 Potholes in the Road to Information Quality" IEEE, vol:30 iss:8 pg:38, 1997.
- [8] W. J. Fernihough. Suggested procedures for generator unit status reporting. Power Apparatus and Systems, IEEE Transactions On PAS-100(6), pp. 2775-2783. 1981.
- [9] J. Paska. Reliability and performance indices of power generating units in poland. Presented at International Conference on Probabilistic Methods Applied to Power Systems, 2004, .
- [10] R. Billinton and J. Ge. A comparison of north american generating unit outage parameters and unavailability indices. Presented at Electrical and Computer Engineering, 2002. IEEE CCECE 2002. Canadian Conference On. 2002, .

New Idea to Estimate the Reliability and Failure Probability with quantitative Methods for Critical Applications

Ossmane Krini

Department of Computer Architecture
and System Programming
University of Kassel,
Germany

krinikrini@web.de

Josef Börcsök

Department of Computer Architecture
and System Programming
University of Kassel,
Germany

ABSTRACT

In order to use electronic systems comprising of software and hardware components in safety related and high safety related applications, it is necessary to meet the Marginal risk numbers required by standards and legislative provisions. Existing processes and mathematical models are used to verify the risk numbers. On the hardware side, various accepted mathematical models, processes, and methods exist to provide the required proof. To this day, however, there are no closed models or mathematical procedures known that allow for a dependable prediction of software reliability. This work presents a method that makes a prognosis on the residual critical error number in software. Conventional models lack this ability and right now, there are no methods that forecast critical errors. The new method will show that an estimate of the residual error number of critical errors in software systems is possible by using a combination of prediction models, a ratio of critical errors, and the total error number. Subsequently, the critical expected value-function at any point in time can be derived from the new solution method, provided the detection rate has been calculated using an appropriate estimation method. Also, the presented method makes it possible to make an estimate on the critical failure rate. The approach is modeled on a real process and therefore describes two essential processes-detection and correction process.

Keywords

Failure rate, reliability, critical faults and estimation.

1. INTRODUCTION

The two processes above are described as two differential equations. The new approach performs a detection of critical errors as well as a prediction after a repair time. If a critical error is detected, a certain amount of time is required to fix it. The time difference also is introduced as a mathematical time-dependent function. This approach provides following advantages and differences to existing methods:

- One aspect is the newly introduced approach which uses one of the conventional reliability models to estimate the total error count and requires real failure data to ultimately make a prediction regarding the critical errors.
- Instead of making a prognosis on critical errors by using already known models, which is often not possible, it suffices to base a better prognosis on a relatively low number of errors using the new approach.

- The currently used software reliability models frequently consist of very unrealistic model assumptions. These model assumptions can neither be used for the critical error prediction or for the error analysis. In the new approach, the repair time will be explicitly described and applied.
- With the repair time, the correction process is described and adjusted to the new method. By doing this, a realistic prediction of critical errors can be conducted.

Increased functionality is identified in current and future software-systems. Statistics and analysis of failures of such complex software-systems reveal that more than half of the system failures are attributed to the software components. Therefore, it is necessary from a scientific and economic standpoint to use reliability models that are capable of predicting software failures. Software reliability is the probability of a failure-free function of a computer program at a specific time in a specific environment. Hardware components are assumed to only fail spontaneously. The failures are due to manufacturing defects, wear, or external factors. Errors in software that cause failures exist statically and are permanent. There are no random malfunctions in software like there are in hardware. The random occurrence of incorrect software behavior usually occurs through the many variations and combinations of software. Today, many programs exist to evaluate and predict hardware reliability. However, even to this day, there are no reliable and meaningful software probability models available to make a general prognosis for safety-critical faults. Since there is no preferable or particular software reliability model, software developers are forced to test a number of individual models. The predictions resulting thereof, however, do not always offer the desired level of accuracy.

2. CONVENTIONAL SOFTWARE RELIABILITY MODELS FOR PREDICTION OF SOFTWARE-ERRORS

Intuitively, a software-system is understood to be reliable when the reasonable user's expectations are met without failures throughout a time period. This time period is represented with the variable. Mathematical software reliability models aim to quantify the reliability. Such models are well founded in theory and are used particularly for safety critical systems. This means that the random variables $T_1, T_2 \dots T_n$ for the time between the failures can still be seen as exponentially distributed, but the parameters of the distribution $\lambda_1, \lambda_2 \dots \lambda_n$ can be different. Software reliability models (SRM) make certain assumptions as to how the

parameters interact. About eighty such models are known in literature [1]. The purpose of such software reliability models is the prognosis of the expected reliability of a system based on data from the past. The future failure rate can be and is predicted from the known times between failures and until a point in time or T . It is assumed that the pure operating time is measured; this means that the accumulation of operating time is stopped as soon as the system is out of operation or in test mode.

The first binomially distributed models were developed by Jelinski and Moranda and were based on birth and death processes. A binomial distributed model has an unknown set number of errors u_0 . Every error has the same probability of occurrence. The hazard rate of every error $z_a(t)$ remains the same over time and is constant (Φ). Furthermore, a binomially distributed error model assumes that the time between failures are independent from each other. If a failure occurs at point in time, it is corrected immediately at point in time t_x . In this approach, it is assumed that an error correction does not lead to new errors. This is why Jelinski and Moranda introduced the following expression for the failure rate:

$${}^s\lambda(t) = c \cdot N_i, \quad t \in [t_{i-1}, t_i]$$

where c is a constant and N_i is the residual error number at point in time t_i . Thus, following can be written with the Equation (1):

$$\begin{aligned} {}^s\lambda(t_1) &= c \cdot N_1 = c \cdot N \\ {}^s\lambda(t_2) &= c \cdot N_2 = c \cdot (N_1 - 1) = c \cdot (N_1 - 1) = c \cdot (N - 2 + 1) \\ {}^s\lambda(t_3) &= c \cdot N_3 = c \cdot (N_2 - 1) = c \cdot (N_1 - 1 - 1) = c \cdot (N - 3 + 1) \\ &\dots\dots\dots \end{aligned}$$

The failure rate of a binomially distributed error model can be defined as step function [1].

$${}^s\lambda(t) = c \cdot (N - i + 1)$$

Software reliability ${}^sR(t)$ is defined according to Musa is as follows:

$${}^sR(t) = e^{-\int_0^t {}^s\lambda(\tau) d\tau} = e^{-\int_0^t c(N-i+1) d\tau} = e^{-c(N-i+1)t}$$

The software failure probability ${}^sF(t)$ thus results for:

$${}^sF(t) = 1 - e^{-\int_0^t {}^s\lambda(\tau) d\tau} = 1 - e^{-\int_0^t c(N-i+1) d\tau} = 1 - e^{-c(N-i+1)t}$$

To make an estimate for the unknown parameters N and c , the maximum-likelihood-method is used. This method determines the parameters so that the mean value function, also called the expectation value, is maximized.

The Poisson distributed Error Model is the best-known model among software reliability models. This model has been developed by Musa at the AT&T Bell Laboratories. Musa was an

important advocate for the use of models and thus, his models are used in various areas. The strengths of this model are in the estimation of test effort in relation to the execution time to achieve a specified MTBF (mean time between failure). $M(t)$ is the number of failures in the time interval from $[0..t]$. $M(t)$ is a Poisson process with mean value function:

$${}^s\mu(t) = E(M(t)) = \sum_{n=0}^{\infty} n \cdot P(M(t) = n)$$

This results to the software failure rate:

$${}^s\mu'(t) = {}^s\lambda(t) = \frac{d}{dt} e(M(t)) = \beta_0 \cdot \beta_1 \cdot e^{(-\beta_1 \cdot t)}$$

and must decrease monotonously. ${}^s\mu''(t) < 0$, for instance, is sufficient. It is also possible that ${}^s\mu'(t)$ deviates. However in this case, at least the envelope curve of the peak values must fall monotonously. When the parameter β_1 becomes too large, the failure rate decreases faster and runs to zero. If, however, the parameter becomes too small, the failure rate decreases slower. In all cases, though, the failure rate function goes towards zero. In practice, the Musa execution model, the sMTTF -value (Mean-Time-To Failure) value, is not determined, but the mean value function ${}^s\mu(t)$ provides the number of errors during a warranty period. Musa makes the following assumptions:

- The failure rate ${}^s\lambda(t)$ is piecewise constant. The result is that the failure rate between two errors is distributed exponentially.
- ${}^s\lambda(t)$ is not proportional to the residual error count and
- there possibly is a non-linear relationship between time scale and the "natural" time.

3. DESCRIPTION OF DETECTION AND CORRECTION PROCESS FOR PREDICTION OF CRITICAL ERRORS IN SOFTWARE-SYSTEMS

Most conventional methods distinguish between error-detection and error-correction. However, to perform a realistic prognosis, the model assumptions are to be modified and adapted. This chapter is going to present a new approach [3]. The detection and correction process for critical errors is mathematically described and fed into a finite as well as an infinite stochastic model. Thus a realistic prediction for the error detection and error elimination will be guaranteed.

One of the known model assumptions is that for every error F_a that occurs at a particular point in time t_a , is immediately corrected at time t_a . Therefore, there is no repair time $\Delta\tau$. Nonetheless, in reality, an error is not corrected immediately after it has been detected. The time that is required to remove a critical error of course depends on other factors including complexity of the detected error. Likewise, the time depends on how quickly the team can reproduce the critical error, find the cause and fix it. Therefore, the repair time during the correction process should not be neglected, figure 1.1.

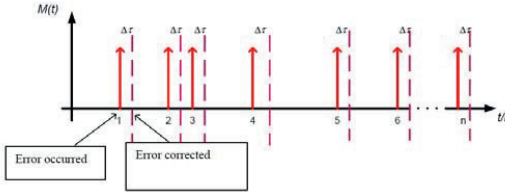


Figure 1.1. Constant Delay-time.

Most conventional models do not consider the repair time. Therefore, these models make inaccurate forecasts, especially regarding critical errors. Hence, the repair time must be considered in the prediction process [3].

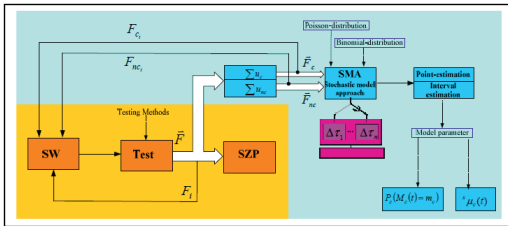


Figure 1.2. New algorithmic process.

In the first part of this chapter, repair time $\Delta\tau$ is mathematically described and assumed to be constant over time. The constant delay time serves to emphasize the difference to conventional methods.

In the second part, however, the repair time $\Delta\tau$ is considered realistically as a time-dependent function $\Delta\tau(t)$. Subsequently, the delay time is fed into different models in order to make a prognosis regarding the residual number of critical errors.

The maximum-likelihood-method is considered for parameter estimation [4].

In order to describe a detection and correction process, different requirements apply. Both a detection and a correction process run in a non-homogenous Poisson process. The critical errors are independent from each other. The mean number of critical errors in the time interval $[t, t + \Delta\tau]$ is proportional to the mean number of the remaining not corrected critical errors. Every critical error that occurs is therefore corrected with no new errors added. The delay time $\Delta\tau$, until the correction process can be assumed, is to be constant. Therefore, the correction process is considered to be a detection process delayed by $\Delta\tau$. The following therefore applies:

$${}^S\mu_{c/nc}(t) = {}^S\mu_{c/nc}(t - \Delta\tau)$$

where ${}^S\mu(t)_c$ is the cumulative function of a critical error and

${}^S\mu(t)_{nc}$ is the cumulative function of a non-critical error.

Figure 1.3 left shows the progression of both functions. It can be seen that the correction curve in blue is shifted by a constant factor $\Delta\tau$.

At point in time τ_x a certain number of errors N is detected, but corrected at point in time τ_c . Since every critical error requires a different amount of time to be corrected, it is very unrealistic that a delay time $\Delta\tau$ can be assumed to be constant [2]. Though, for a new approach, a nonconstant delay time $\Delta\tau(t)$ is chosen, figure 1.3 right.

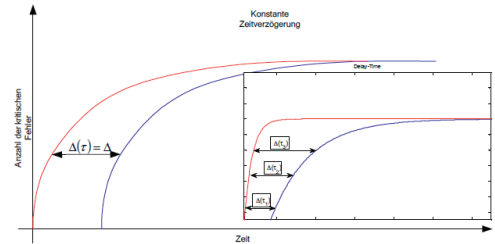


Figure 1.3. Different mean value functions (Delay-time).

There is no constant delay time $\Delta\tau(t)$ between the detection and correction processes. Now it remains to be clarified how the delay time $\Delta\tau(t)$ is described mathematically. The delay time $\Delta\tau(t)$ can be described at will; however, the delay time depends on the software-system that is going to be analyzed. Hence, the time shift $\Delta\tau(t)$, for instance, can be assumed to be linear:

$$\Delta\tau(t) = t$$

This would mean, however, that proportionality exists between the abscissa, the time axis, and the ordinate, which is the number of errors. Because of that, a linear delay time $\Delta\tau(t)$ is not recommended for this approach.

Another consideration, which arose during the process of this thesis, is to consider the delay time as an exponential function with a negative exponent.

$$\Delta\tau(t) = a \cdot e^{-t}$$

The goal is to obtain a cumulative function as a result for the correction process that reflects a relative approximation of reality. Since the “In-func” can only adopt positive values, another promising solution seems to be the selection of a logarithmic function. This could lead to a so-called “learning process”. To ultimately examine the effect of the exponential or logarithmic delay time in the correction process, the detection and correction process must be described mathematically.

The following first order differential equation applies for the critical detection process:

$$\frac{\partial {}^S\mu_c(t)}{\partial t} = d_\psi(t) \cdot \left(\sum_{i=0}^n u_{0,i} - {}^S\mu_c(t) \right)$$

Function $d\psi(t)$ describes the detection rate for the critical errors and shall be assumed to be constant in this work. Index “ ψ ” indicates that the detection rate refers to the new approach and the newly introduced constant Ψ . Ψ represents the ratio between the critical error and the total number of errors:

$$\psi = \frac{\sum_{i=0}^n u_{0_{c_i}}}{\sum_{i=0}^n u_{0_{c_i}} + \sum_{i=0}^n u_{0_{nc_i}}} \begin{cases} u_{0_{c_i}} \Rightarrow & \text{critical number of faults} \\ u_{0_{nc_i}} \Rightarrow & \text{no critical number of faults} \end{cases}$$

The rate $d\psi(t)$ is to be always seen as unknown and therefore needs to be estimated by using the maximum-likelihood-method. The detection process runs a non-homogeneous Poisson form. Furthermore, all critical errors shall be independent from each

other. $\sum_{i=0}^n u_{0_{c_i}}$ represents the total number of critical errors.

$\overset{S}{D}\mu_c(t)$ describes the expectation value of the critical errors in the detection process. According to the “postulate” at the beginning of the prognosis $t=0$, there are no critical errors. Because of this, the side condition follows the differential condition [2]:

$$\overset{S}{D}\mu_c(0) = 0$$

Thus, the inhomogeneous differential equation of the first order can be solved completely.

$$\frac{\partial \overset{S}{D}\mu_c(t)}{\partial t} = d_\psi \sum_{i=0}^n u_{0_{c_i}} - d_\psi \overset{S}{D}\mu_c(t)$$

$$\frac{\partial \overset{S}{D}\mu_c(t)}{\partial t} + d_\psi \overset{S}{D}\mu_c(t) = d_\psi \sum_{i=0}^n u_{0_{c_i}}$$

The delay time $\Delta\tau(t)$ is irrelevant, because, although the critical error has been detected, it has yet to be corrected. In the most cases, the error cannot be corrected immediately because the system is in operation.

Now the correcting process of the critical error is described mathematically as follows:

$$\frac{\partial \overset{S}{C}\mu_c(t)}{\partial t} = c_\psi(t) \cdot \left(\overset{S}{D}\mu_c(t) - \overset{S}{C}\mu_c(t) \right)$$

$\overset{S}{C}\mu_c(t)$ describes the expectation value of the critical errors. The correction rate can also be interpreted using Equation (16):

$$c_\psi(t) = \frac{\frac{\partial \overset{S}{C}\mu_c(t)}{\partial t}}{\left(\overset{S}{D}\mu_c(t) - \overset{S}{C}\mu_c(t) \right)} = \frac{\overset{S}{C}\mu_c'(t)}{\overset{S}{D}\mu_c(t) - \overset{S}{C}\mu_c(t)}$$

Equation (17) shows, that the correction rate $c_\psi(t)$ can be seen as the error correction rate per detected errors, but as not corrected “critical” errors. In reality, the correction rate depends on the complexity of the problem being analyzed, the abilities of the test team, and the time restrictions for the handover of the finished software to the customer.

Equation (17) can now be solved with the following side condition:

$$\overset{S}{C}\mu_c(0) = 0$$

The solution of the differential equation for the correction process then leads to the following expectation value:

$$\overset{S}{C}\mu_c(t) = e^{-c(t)} \cdot \left(\int_0^t \left[\sum_{i=0}^n u_{0_{c_i}} + \sum_{i=0}^n u_{0_{nc_i}} \right] \cdot c(s) \cdot e^{c(s)} \cdot (1 - e^{-D(t)}) \right] ds$$

where:

$$D(t) = \int_0^t d(s) ds$$

$$C(t) = \int_0^t c(s) ds$$

In order to maintain the mathematical overview and for the sake of simplicity, the correction rate $c_\psi(t)$ shall also be assumed to be time-independent in the scope of this work. When the differential equation is now solved, the following expectation value is obtained for the correction process:

$$\overset{S}{C}\mu_c(t) = \left(\sum_{i=0}^n u_{0_{c_i}} \right) \cdot \left(1 - (1 + c_\psi \cdot t) \cdot e^{-c_\psi \cdot t} \right)$$

$$\overset{S}{C}\mu_c(t) = \left(\sum_{i=0}^n u_{0_{c_i}} \right) \cdot \left(1 - e^{-c_\psi \cdot t \cdot \frac{\ln(1+c_\psi \cdot t)}{c_\psi}} \right)$$

Consequently, the delay time $\Delta\tau(t)$ of the correction process can be seen explicitly from the Equation.

4. PREDICTION AND CALCULATION BASED ON A FINITE STOCHASTIC MODEL WITH THE REAL FAILURE DATA

The finite stochastic model is very similar to the finite binomial distributed model of section 2, but has, due to the newly introduced correction processes, the most important property of continuous improvement of the critical error rate over time. The stochastic model is used for the determination of critical error distribution over a particular time period. Furthermore, the model serves as an estimation of the number of remaining critical errors [2]. The finite-stochastic model that is based on the correction process serves as an estimate of the additional time required for the improvement of a specific reliability.

Using the equations mentioned above for the detection and correction processes, interesting expectation values are obtained [2].

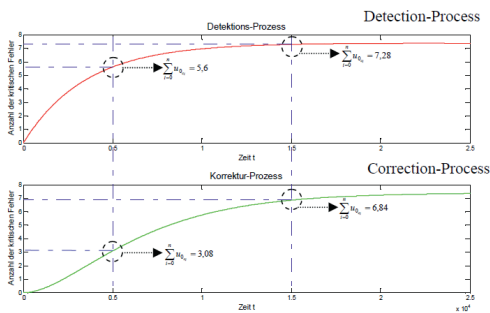


Figure 1.4. Detection- und Correction-Process.

Figure 1.4 shows a realistic and time curve of the detection and correction process. This process is based on real failure data. The effect of the time delay $\Delta \tau(t)$ can be seen clearly in the correction process. During the so-called "learning phase" a critical number

$$\sum_{i=0}^n u_{0_{ci}} = 5,6 \text{ are detected at a certain point in time}$$

$$t=0,5 \cdot 10^4 \text{ h, } \sum_{i=0}^n u_{0_{ci}} = 7,28, \text{ critical errors are detected. At}$$

$$\text{the same time } \sum_{i=0}^n u_{0_{ci}} = 6,84, \text{ critical errors are corrected.}$$

A realistic prognosis is obtained with the introduction of correction processes for critical errors. Conventional models cannot predict critical errors and secondly the models do not

provide correction processes. Figure 1.5 shows conventional models on the left. It can be seen that neither the detection process nor the correction processes provide meaningful information. The conventional correction process fails, because almost zero errors have been corrected at time $t=0,5 \cdot 10^4 \text{ h}$ (figure 1.5, picture 3). Even with the time progressing to time $t=2,5 \cdot 10^4 \text{ h}$, no satisfying prognosis can be achieved.

The newly introduced method in figure 1.5, right, establishes a more reliable prognosis at time $t=0,5 \cdot 10^4 \text{ h}$.

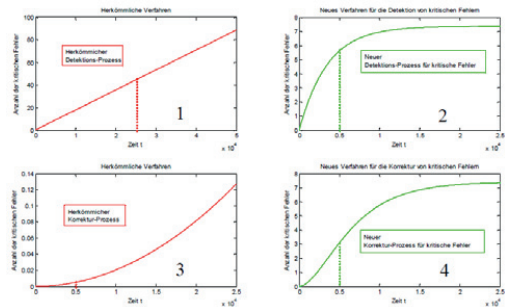


Figure 1.5. Detection- und Correction-Process.

5. SUMMARY

It can be noted that conventional modes cannot describe a correction process for critical errors since only minimal failure data is available. Every convention model requires far more than the minimal amount of failure data to make reliable forecasts. This work presents a method that performs a prognosis of critical errors based on a correction process. The newly established correction process, however, requires that all errors are corrected completely.

6. REFERENCES

- [1] Musa, J.D.: Software Reliability Engineering, McGraw-Hill, New York, 1998.
- [2] Huan-Jyh Shyur: A stochastic software reliability model with imperfect-debugging and changepoint", The Journal of systems and software, 2003.
- [3] O.Krini, J.Börcsök: Different Approaches to Predict Software Reliability, Proceedings of 24th international congress, 2011.
- [4] A. Birolini: Reliability Engineering: Theory and Practice", Springer Verlag, Heidelberg, 2007.
- [5] Lyu, M.: "Handbook of Software Reliability", McGraw-Hill, New York, 1996

Modified Physics of Failure approach for Reliability prediction of Electronic Components

Adithya Thaduri^{#S1}, A K Verma^{&2}, V Gopika^{*3}, Rajesh Gopinath^{*4}, Uday Kumar^{S5}

Department of Electrical Engineering, IIT Bombay, Powai, Mumbai, India

1adithya.thaduri@gmail.com

& Stord/Haugesund University College, Haugesund, Norway

2akvmanas@gmail.com

*RSD, BARC, Trombay, Bombay, India

3vgopika@barc.gov.in

4rajeshgopin@gmail.com

SDivision of Operation and Maintenance, Lulea University of Technology, Lulea, Sweden

5Uday.Kumar@ltu.se

ABSTRACT

Prediction of reliability is essentially required for design for reliability, warranty periods, life cycle costs and maintenance with prior to the installation of components. Reliability prediction of electronics advances from an era of military standard books, Telcordia, PRISM, Bell Core etc. to the physics of failure approach. Most of the limitations of constant failure methods are being encounter by the PoF approach due to its advanced methodology of finding failure mechanisms by root cause of failure analysis. But this approach has its own challenges as it requires detailed information on materials, process, technology and other specifications and at most of the times this data is confidential from the industries. On the other hand, probability and statistics methods provide quantitative data with reliability indices from testing by experimentation and by simulations. In this paper, qualitative data from PoF approach and quantitative data from the statistical analysis is

combined to form a modified physics of failure approach. This methodology overcomes some of the challenges faced by PoF approach as it involves detailed analysis of stress factors, data modeling and prediction. A decision support system is added to this approach to choose the best option from different failure data models, failure mechanisms, failure criteria and other factors.

KEYWORDS

Physics of Failure, Probability and Statistics, Experimentation, Simulation, Decision Support System

1. INTRODUCTION

Reliability is one of the important factors in the operation of a device and miscalculation of these indices results in various losses like repair/replacement costs, human factors, time and reputation. Hence, efficient reliability prediction is needed before the installation of the components

and appropriate changes can be made at the design stages. Conventional reliability prediction methods such as Mil-Hdbk 217F, Telcordia, Bell Core, PRISM etc implements constant failure methods and believed to be true in the era of 1980s and 1990s. But due to the advancements in the technology, these methods are no longer adequate to define the characteristics as there are so much variability in the design and fabrication of devices. Especially, electronics spreads out rapid developments in all the aspects and applicable to everywhere and even replaced the other technologies due to its control aspects, miniature and cost effectiveness. Some of the devices are used in safety, security and military areas where the availability is the major concern and incorrect operation leads to the unsafe shutdown. Moreover reliability aspects and prediction is critical to these components and this paper provides advanced physics of failure methodology for finding failure characteristics and reliability indices. The following Table 1 demonstrates various traditional prediction methods the differences between the values of time to failures of DC-DC converter constraints the ambiguity and risk in selecting appropriate figure.

Table 1: Comparisons of different reliability prediction models [1]

Reliability Prediction Model ¹	Company	1 Watt DC-DC Converter ²					
		25°C		85°C		100W AC-DC PSU ³	
		Hours	Years ⁴	Hours	Years ⁴	Hours	Years ⁴
MIL-HDBK-217F (EXAR 7.0)	A	31,596,574	3606.9			686,771	78.4
MIL-HDBK-217F Notice 2	B	832,000	95.0	86,000	9.8		
MIL-HDBK-217F Notice 1	C	156,000	17.8	124,000	14.2		
Telcordia SR332 Parts count	D	89,380,000	10203.2	29,260,000	3340.2		
Telcordia SR332 Parts stress	D	104,200,000	11895.0	57,160,000	6525.1		
Siemens SN29500 (IEC61709)	A	80,976,217	9244.1			1,554,055	177.4
IRIS5 Parts stress	B	2,465,000	281.4	849,000	96.9		
HRD4 Parts count	B	1,132,000	129.2	1,132,000	129.2		
MIL-HDBK-217F (EXAR 7.0)	A	31,596,574	3606.9			686,771	78.4
Telcordia SR332 Parts count	E					1,418,000	162.0

Physics of failure prediction methodology lay emphasis on the root cause of failure following fundamentals of physics of materials considered as white box testing. Electronic devices were fabricated of different materials like Silicon, Germanium, Copper, Aluminium etc., with predefined technology and processes like deposition, etching and masking under controlled environment. Most of failures happened due to changes in thermal activation, changes in defects of molecules, activation energy in the materials and correspondingly there was degradation in the technology and performance factors. Depends on

the failure characteristics, several failure mechanisms were modeled and categorized according to the cause, material, and failure point location etc., in the literature. As from the application and field environment, appropriate method was selected, analyzed and TTF was calculated accordingly. There were significant advantages to this methodology like reliability design, condition monitoring, improvement in LCC and component selection to the application involved. This method requires sophisticated tools for failure analysis and advanced tools for analyzing the simulated data. Still, this methodology also has challenges like insufficient data from the manufacturer, needs expert judgment and also time taking process [2, 3, 5, 6].

On the other hand, statistical methods were widely available in order to find out the reliability indices from the test data. This method was also considered as black box testing which concentrate on available data and proper model was selected depends on the application. There were possibilities to analyze the data and generated model to extract enormous amount of information to characterize the performance parameters. Some of them include design of experiments, accelerated testing, regression analysis, etc. Even, there were several tools available for model selection, mathematical formulation and model analysis. This methodology has some advantages like time consuming, no need for manufacturer data and parameter analysis.

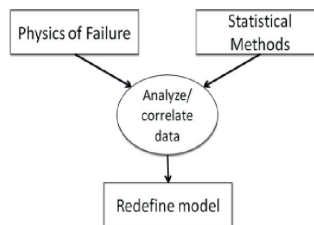


Figure 1: Short idea of Modified approach

Inclusion of multidisciplinary science and engineering approaches was very effective in solving of real life problems and our modified approach was combination of both physics of failure (deterministic) and statistical (probabilistic) approaches in Figure 1. This advancement

methodology first starts with the proper understanding of basic failure physics of the component and process the physics of failure methodology. This knowledge was fed to the statistical approach to further refining of data for accurate models. Finally, we get three faces of models; history and literature, white box and black box models and these were sent to decision support system. The other inputs to this system were life cycle costs and regulatory requirements.

2. BLOCK DIAGRAM

The proposed methodology was divided to five stages; component description and history, literature failure survey, experimentation and simulation and decision methodology. This was represented as in fig 2 and all of them are interacted to each other. A short description of these blocks was discussed below.

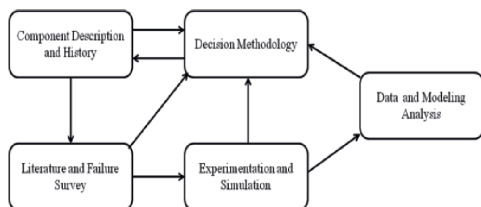


Fig 2: The short notation of modified physics of failure approach

2.1 Component Description

The accuracy of physics of failure methodology depends on the abundance of the information required to qualify and quantify the failure phenomenon for the respective component. Hence, hefty amount of information is required on all sides to get best out of the prediction.

2.2 Literature and Failure survey

As part of information, investigation on the component from various sides to acknowledge the failure was needed for implementation. Research was to be carried out by considering the compelling stress parameters that could precipitate in the field in correspondence with the failure mechanisms, failure modes and degradation analysis associated with it. In order to verify the failures, respective method of failure analysis was selected.

2.3 Experimentation and Simulation

The information acquired from above two stages was helpful in designing the experiments to be carried out on the component for hypothesis testing. This stage involves experiment testing using the PCBs and possible simulations using advanced tools which can be carried out simultaneously. The main aim of this stage is to acquire large amount of data for analysis and modeling.

2.4 Data and Modeling analysis

In this stage, data was converted into information required for processing and modeling the failure characteristics of the component. This stage requires statistical data analysis and tools to characterize each parameter and postulate some of the important results to define failure. This data was further essential for developing models using different methods.

2.5 Decision Methodology

This is the important stage in defining the failure prediction of the component. This stage capitalize the information from all the stages and also several other non-technical issues to decide on which alternate was to be selected for the reliability growth of the component. This stage was an advanced multi-objective optimization of several factors that provide different solutions. The job was to discuss and analyze alternate solutions and decision can be made.

The advanced modified physics of failure methodology considering the above stages was shown in fig 3. The description of each of the blocks was demonstrated below.

3. DATA FLOW

Initially, the component was described thoroughly to get enough failure information. First need to check whether that component was existed in the field, and if it's available similar item analysis and if failures were present an extensive methodology was carried out and correspondingly failure analysis, failure mechanism and failure modeling was implemented to get an idea of the component. We need also to check whether the component was analyzed in the literature that information was also

stored. After an extensive research and inputs from the similar and failure analysis, a detailed methodology needs to be planned in the sequential order of failure modeling, experimentation, simulations and statistical and data modeling. According to the plan, everything was executed simultaneously to reduce the amount of time in testing. After getting the data, several analyses of

factors was conducted and modeling was developed from various methods. The essential information from all the blocks were given as inputs to the decision support system where it provides the best alternative was selected and considered as technique for reliability growth. This information was stored in the component database where it was useful for further analysis.

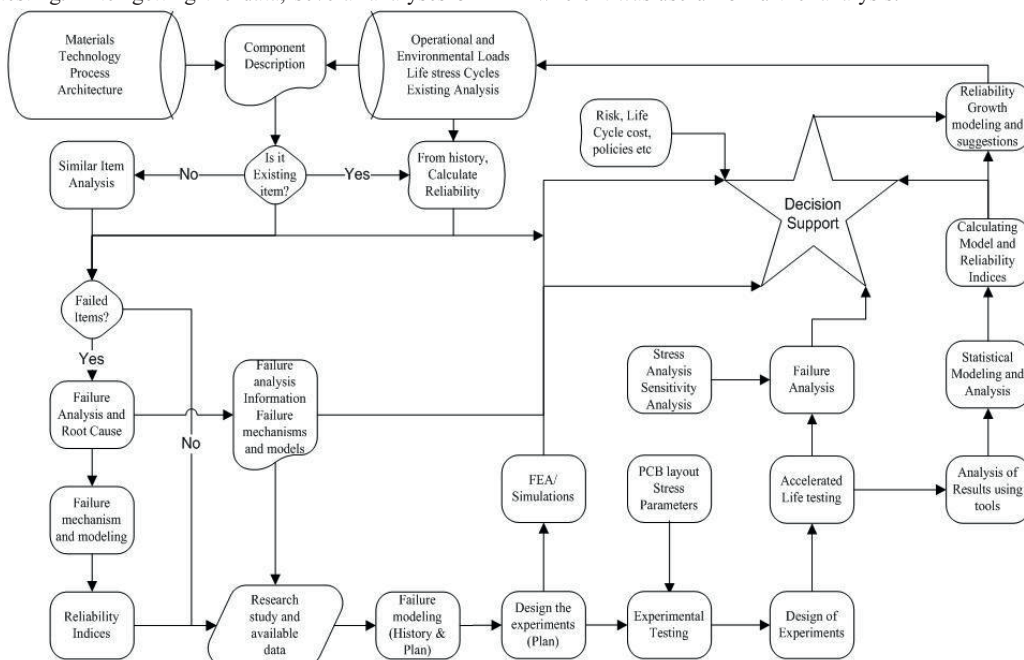


Fig 3: Advanced block diagram of proposed physics of failure approach.

3.1 Component Description

As informed above, this analysis requires as much as information for the pre- and post processing examination. Hence, the component was collected from the various data and sources are essential in building up data [4]. The resources required for data part are:

- i. Materials used for fabrication and its properties
- ii. Diagrams for layout of internal chip structure
- iii. Various stresses effecting at the field and its performance
- iv. Architecture used for design

- v. Processes carried out during the fabrication
 - vi. Design of the circuitry
 - vii. and technology implemented for fabrication
- The resources required for data part are:

- i. Manufacturer of the product/item
- ii. Consumer data supplied
- iii. Similar items that was earlier carried out in house
- iv. Manuals for that component
- v. Field information
- vi. And design team for information

3.2 Literature and History Data

As for the failure study, learning the literature was necessary for understanding the behavior of the component under the failure considerations [4].

The aspects need to be considered in literature are:

- i. Stress parameters in and off the field
- ii. Reliability growth techniques available
- iii. Testing information and setup
- iv. Possible failure point locations (weak areas)
- v. Failure modeling methods and techniques and failure criteria
- vi. Failure analysis using sophisticated equipments
- vii. Failure mechanisms that effect the behavior of performance parameters
- viii. Operational life cycle of the component

The aspects to be considered were provided as:

- i. Field information
- ii. Prediction of life using MilHdbk and other standard handbooks
- iii. Reliability indices to be considered
- iv. Datasheets from the manufacturer
- v. Failure data provided in the research

3.3 Similar Item Analysis

Several techniques have been developed and used in performing very early predictions of item reliability before any characteristics of the system design have been established [4].

- i. Defining the new item
- ii. Identifying an existing item with nearly comparison
- iii. Obtaining and analyzing historical data
- iv. Drawing conclusions on the level of reliability

Major factors for a direct comparison of similar items should include: Item physical and performance comparison, design similarity, manufacturing similarity, similarity of the service use profile, program and project similarity and proof of reliability achievement.

3.4 Reliability Indices

There are several indices are present to define reliability of the component. They are time to failure, failure rate, percentage of degradation and probability. An appropriate parameter was selected by limiting with the failure criteria of the component. It comes under one of the parameters in the design considerations.

3.5 Failure Analysis

Failure analysis consists of confirming reported failures and clarifying failure modes or mechanisms using electrical measurements and various scientific analysis technologies. This section introduces specific failure analysis methods. However, before performing the actual analysis work it is necessary to thoroughly investigate failure circumstances and accurately understand the failure contents. This makes it possible to determine the optimum analysis methods and carry out swift processing.

As semiconductor devices become more highly integrated and incorporate more advanced functions, manufacturing processes are becoming more miniaturized and complex, and include diverse reliability factors. In addition, semiconductor devices have come to be used over an extremely wide range of fields, so failure causes and mechanisms are also complex. Under these circumstances, an extremely high reliability level is required of semiconductor devices. Reliability must be built in from the device development stage to the manufacturing stage in order to ensure a high level of reliability.

There are several destructive and non-destructive sophisticated methods are available at several handbooks and simulations in order to characterize the device at various levels, to implement failure analysis and also to find failure point location [2, 3, 5, 9]. This is the comprehensive list of several non destructive failure analysis techniques applicable for each failure mechanism.

- Hot carrier injection: hot spot: photo emission analysis, thermal analysis, SEM, Liquid Crystal method
- TDDB: oscilloscope for detection of breakdown voltage.
- Electromigration: Electron Probe Micro analysis
- To quantify the internal Image: Image Analyzing System
- Temperature and heat related failures: Thermal Analysis System
- Impurities like S, P, F, Cl, Br and I: X-Ray Fluorescence Spectrometer and also FTIR
- Corrosion: Time of flight secondary ion mass spectrometer.

- ESD: Optical beam induced current analysis (OBIC), TEM, Optical Microscope, SEM
- Latchup: Optical beam induced current analysis (OBIC)
- For Electric Measurements: IC Tester, Oscilloscope and Curve Tracer
- Surface Analysis: Transmission Electron Microscope (TEM)

3.6 Failure Mechanisms

Advanced integrated circuits (ICs) are very complex, both in terms of their design and in their usage of many dissimilar materials (semiconductors, insulators, metals, plastic molding compounds, etc.). For cost reductions per device and improved performance, scaling of device geometries has played a critically important role in the success of semiconductors. This scaling—where device geometries are generally reduced by 0.7x for each new technology node and tend to conform to Moore’s Law —has caused the electric fields in the materials to rise (bringing the materials ever closer to their breakdown strength) and current densities in the metallization to rise causing electromigration (EM) concerns. The higher electric fields can accelerate reliability issues such as: time-dependent dielectric breakdown (TDDB), hot-carrier injection (HCI), and negative-bias temperature instability (NBTI). This failure mechanisms behave differently depends on the technology such as CMOS, BJT and other semiconductors, process, manufacturer etc [2, 3, 5]. In addition, the use of dissimilar materials in a chip and in the assembly process produces a number of thermal expansion mismatches which can drive large thermo-mechanical stresses. These thermo-mechanical stresses can result in failure mechanisms such as stress migration (SM), creep, fatigue, cracking, delaminating interfaces, etc. [6, 7, 8]

3.7 Failure Modeling

In order to predict the life time of the component, an appropriate model was designed or developed or selected which depends on the data generated from the experimental and simulation results. Apart from the standard physics of failure models, several models that were generated from the

statistical results were also compared to define behavior of the stress and performance parameters [7, 8, 12]. As mentioned in Fig 4, the model depends on the field and testing data, failure mechanisms and modes, stress parameters involved and by reference as failure criteria; it can be compared with the existing models.

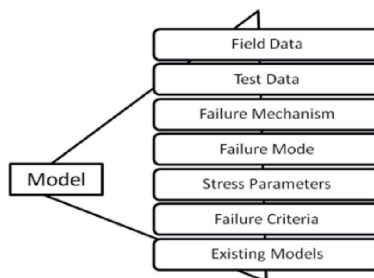


Fig 4: Model dependence parameters

3.8 Design of Experiments

This technique was well established technique to find the variability of the input stress parameters and its effect on the performance parameter. Design of Experiments (DOE) techniques enables the designers and fabrication engineers to determine simultaneously the individual and interactive effects of many stress factors with respective levels that could affect the output results in any design [13]. DOE also provides a full insight of interaction between parameters and thus efficient in converting standard design into a robust one. DOE helps to make concentrate on the sensitive stress-levels and sensitive areas in designs that cause problems in degradation, best performance and yield. Designers are then capable to reconfigure these parameters to reduce problems and correspondingly produce robust and higher designs before production. Design of experiments (DOE) is the design of any information-gathering exercises where variation is present, whether under the full control of the experimenter or not. Stress factors, levels and their interactions are tabulated for response curve and provides and runs that will best and worst solutions. In standard procedure, Taguchi method was implemented by considering the stress factors with levels with some number of runs. In general, there was a risk in selecting in levels of parameters.

In our work, we modified the conventional DOE into two steps: screening step and testing step. Initially appropriate samples were selected for each stage for repeatability and accuracy. The first step demonstrates the observance of input stress parameters on the output parameter. The response curve generated from this step provides the increase/decrease of respective stress parameter results in the degradation of performance output parameter. Then in accordingly the worst levels of the stress was selected for second testing step in constraints with the datasheet of the component. In the testing step, the experiments were conducted from the inputs of step 1. By this methodology, the ambiguity and risk in the selecting the stress levels was eliminated.

3.9 PCB Design and Layout

In order the experiment the electronic component, an appropriate circuit was designed and fabricated using Printed Circuit Board. There were several tools available to design the circuitry to compatible with PCB. The board layout was properly designed to reduce the interspatial effects, size and interoperability. As the experiment was needed to be exposed under stressed accelerated testing, the circuitry need to be designed in such a way that the component under stress was segregated with the other control and power circuitry. This technique helps to reduce the effect the trace changes of other components such as resistor, capacitor and other miscellaneous components on the measured parameters as this components may vary their parameters in according with stress.

3.10 Experimental Testing

After developing the circuit, the items were subjected to the stresses and monitor the output variables using various instruments. The experimental setup consists of various instruments such as voltage suppliers, oscilloscopes, voltage and current meters etc [11]. Accordingly, it was properly maintained in controlled environments to reduce the external noises. As it was needed to be subjected to the accelerated testing, the experiment stage needs to be properly monitored periodically for the effective control. The following stress parameters are temperature, voltage, current, radiation exposure etc. The planned Design of

Experiments was subsequently applied on this circuitry to find the results.

3.11 Simulations

The simulation tools present a virtual environment and also gather information of the respective dimensions by graphical illustration [2]. Simulations are carried out using advanced softwares tools such as Cadence, SPICE, etc by providing inputs of stress parameters, device parameters and limits. This step will run simultaneously with the experimental testing for purpose of comparison with results from experimentation. Finite Element Analysis tools such as Ansys, Comsol, nanoHUB etc are also carried to study the behavior of device and material characteristics.

3.12 Accelerated Testing

In normal operating conditions, the component takes more amount of time to degrade and subsequently results in failure. In order to speed up the testing time, the applied parameters need to be stressed and correspondingly the testing time was reduced [14]. Then using extrapolation and considering the acceleration factor, the failure time at operating conditions was calculated. Hence, accelerated life testing involves acceleration of failures with the single purpose of quantification of the life characteristics of the product at normal use conditions. In the most of the electronic components, the failure time was quite high and hence more rigorous stress levels need to be considered. Accelerating factors and stressed applied, either singly or in combination, include

- i. More frequent power cycling
- ii. Higher vibration levels
- iii. High humidity
- iv. More severe temperature cycling
- v. Higher temperatures

Most common model for temperature is Arrhenius model

$$AF = e^{\frac{E_a}{k}(\frac{1}{T_1} - \frac{1}{T_2})} \quad t_f = Ae^{\frac{E_a}{kT}}$$

where AF = acceleration factor, Ea = activation energy, k = Boltzmann constant, T1, T2 = operating and stress temperatures, tf = TTF

From the normal and operating temperatures, acceleration factor was calculated by substituting this value, time to failure was calculated.

3.13 Analysis of Results

The data generated from both experimentation and simulation was fed to this step. This step involves the behavior study of input stress parameters, design parameters, model parameters with respect to the performance and failure criteria. Individual graphs were also drawn to make some conclusions on the performance. It's like pre-processing stage to characterize the interdependence of the variables and observe the phenomenon of the imminent illustrations. The results were properly analyzed using some of the advanced statistical methods and tools to acquire essential information for further processing.

3.14 Stress and Sensitivity Analysis

This is pre-processing step for failure analysis which provides the affect of stress inputs on the variability of material characteristics using simulations and sensitivity data. This analysis is sub-section of failure analysis in which after acquiring information from the non-destructive testing techniques and simulation data, each and every stress parameter was demonstrated using contour graphs and 3D modeling information. This analysis provides parameters affecting the performance of the component. The sensitivity part provides the interaction between variability of each stress with the output variable.

3.15 Statistical Modeling and Data Analysis

The preprocessing data was applied in this stage to qualify and quantify the data to assess the information. Using some of the statistical methods such as regression, response surface regression, parametric analysis, DOE, quality methods, reliability/survival analysis, accelerated life testing, and support vector machine and other techniques to model the input-output interactions by illustrating the several graphical analysis were generated. This extensive examination of the parameters provides enormous amount of information at which we can judge the performance of the component. The

models generated in the stage were considered as basis for the next steps as it decides the reliability growth techniques. The consideration and analysis the physics of failure models was also taken into account and further modify these models in accordance to the customized design.

3.16 Reliability Indices

From the selected reliability indices at the planning stage, these figures were calculated using developed models such as Physics of Failure, MilHdbk standard handbooks, Response Surface Regression, other regression techniques and support vector machine. All these figures need to be calculated in consideration with the failure criteria. These figures were further compared in a common platform to assess the variability and degradation of the performance parameters with the operating conditions. The outputs of this stage are reliability indices, design range and metrics, safety limits and best parameters for maximum performance.

3.17 Reliability Growth

The final objective of this overall methodology is to find the best design and manufacture alternatives to increase the life time of the component. The techniques required for enhancement in TTF and reduction in degradation of parameters is called reliability growth [10]. This step provides only the prediction so such that uncertainty and confidence levels were also included. The possible reliability growth techniques cover in

- i. Changes in design parameters
- ii. Incorporation of additional circuitry
- iii. Selection of different manufacturer
- iv. Failure site improvement
- v. Fabrication suggestions to manufacture for in-house components.

3.18 Non-technical factors

In deciding the optimal characteristics of the component, several other factors need to be considered at the managerial level. These include risk analysis, government policies, management choices, availability, life cycle cost, human interaction etc. to be considered.

3.19 Decision Support System

This is the final stage of the entire proposed modified block diagram which involves much more productive decision can be made by the information gathered from different parts of the Fig 5. The following figure demonstrates the various factors required as an inputs to the decision support system to finalize the judgment on the component for reliability growth and further to take necessary measures. The inputs to the system are:

- a. Failure Analysis: From acquiring the information of failure point locations at different parts of the block diagram, such as similar item analysis, tested failure analysis, in the literature and from historical data, a final conclusion needs to be stated as input to the support system. This was considered as quality input.
- b. Statistical Models: Models were generated at different parts of the diagram such as in the literature, historical data, failed items and the tested data. An appropriate prediction model was selected for quantitative analysis and thus decisive finding was fed to system.
- c. Simulations: Simultaneously we carried out simulations on the component to identify the stress behavior on the performance parameters and any other essential information was provided to the central system
- d. Risk: The possible risk associated with each alternative was considered as input.

- e. Life Cycle Cost: As cost was one of the main criteria for a business, total cost accumulated for each alternative was considered.
- f. Non-technical factors: Other non-technical factors were also discussed

An expert group consists of reliability engineers, electronic design and fabrication engineers, material engineers, statisticians, field engineers and management need to be discussed on the several alternatives and appropriate solution was to be selected by optimal suggestions from all the people in the group. Each alternative was excessively discussed and generates report considering all the factors and this information will feed back to the database of the component in which this information is useful in further analysis.

4. PREDICTED OUTCOMES

By implementing this advanced methodology, the following productive outcomes provides efficient information as

- Root cause analysis provides the exact failure site location which provides pin pointed improvement area.
- Suggesting different alternatives for the enhancement in reliability
- Reduction in the repair/recall/replacement cost
- Feasible for flexible reliability design using the data w.r.t the application
- Also available for similar item analysis

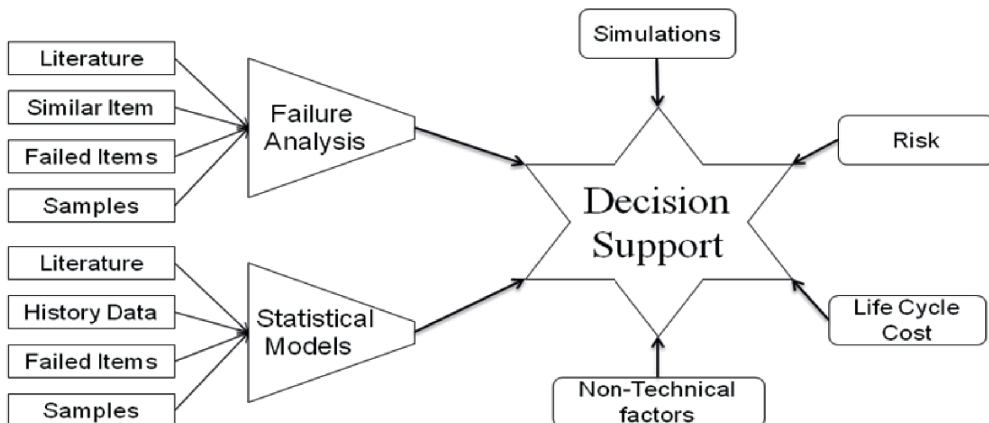


Fig 5: Decision Support System

5. ADVANTAGES

The advantages by using this methodology are

- Proper learning of failures so that future product development, design, strategy and implementation will be more successful.
- Reputation in market due to reliable product outcomes
- Cost, time and human work for recalling, repairing and replacement decreases
- Qualitative and quantitative data is available for the selected component and consider as a basis for advance in design with less time
- Modeling the component as per requirement and provides in-house research
- Increase in time to market depends on supply of products

6. CHALLENGES

This methodology has following challenges and limitations

- Materials, process and technologies are always not available to the customer datasheet by companies due to confidentiality
- Requires more sophisticated instruments (also cost) for analysis which are always not possible.
- Modeling of the failure criteria/degradation phenomena of new materials needs insightful research.
- It takes time to carry out and require cost for all analysis.
- Need expert reviews on the cause of failure

7. CONCLUSION

Physics of failure methodology alone does not provide enough information on the component and hence incorporation of statistical methods will improve the effectiveness of the prediction of the reliability indices. The proposed modified approach accommodates enormous amount of information which also provides several other alternatives which improves the mechanism. But this method is only applicable to the critical parts and components which is very important and provide safety to the costly equipment. This type of rigorous analysis does not require for less important components.

8. REFERENCES

- [1]. "European power supply manufacturers association", MTBF Report, June 2005.
- [2]. Mark White, "Microelectronics reliability: Physics-of-failure based modeling and lifetime evaluation", JPL Publication 08-05, NASA, Feb 2008.
- [3]. Milton Ohring, "Reliability and Failure of Electronic Materials and Devices", Academic Press, 1998.
- [4]. Military Handbook, Electronic Reliability Design Handbook, MIL-HDBK-338B.
- [5]. "Panasonic Reliability of Semiconductor Devices", Panasonic Corporation, April 2000
- [6]. "Renesas Semiconductor Reliability Handbook", Renesas Technology Corp., Nov 28, 2008.
- [7]. "Failure Mechanisms and Models for Semiconductor Devices", *JEP122E*, JEDEC Publication, March 2009
- [8]. "Semiconductor Device Reliability Failure Models", International SEMATECH, May 31, 2000.
- [9]. Perry L Martin, "Electronic Failure Analysis Handbook, Techniques and Applications for Electronic and Electrical Packages, Components, and Assemblies", McGraw Hill, 1999.
- [10]. Ed Habtour Gary S. Drake, and Geetha V. Chary, "Improving the Reliability in the Next Generation of US Army Platforms through Physics of Failure Analysis", *Journal of Failure Analysis and Prevention*, Springer, Vol 2, Number 1, 2012.
- [11]. "EIAJ ED-4701 Semiconductor Device Environment and Durability Testing Methods", Electronic Industries Association of Japan (1994).
- [12]. J W McPherson, "Reliability Physics and Engineering: Time-To-Failure Modeling", Springer, 2010.
- [13]. Lloyd W. Condra, "Reliability improvement with design of experiments", 2nd edn, Marcel Dekker, 2001.
- [14]. Wayne B Nelson, "Accelerated Testing: Statistical Models, Test Plans, and Data Analysis (*Wiley Series in Probability and Statistics*)", John Wiley & sons, 2004.

eMaintenance Ontologies and Data Production

Mustafa Aljumaili
Luleå University of Technology
SE-971 87 Luleå, Sweden
+46 (0)920 49 15 53
Mustafa.Aljumaili@ltu.se

Karina Wandt
Luleå University of Technology
SE-971 87 Luleå, Sweden
+46 (0)920 49 24 83
Karina.Wandt@ltu.se

Ramin Karim
Luleå University of Technology
SE-971 87 Luleå, Sweden
+46 (0)920 49 23 44
Ramin.Karim@ltu.se

ABSTRACT

The demands on the integrity and interoperability of data in maintenance are constantly increasing. For many enterprises, achieving these demands in their operation and maintenance processes is challenging. In order to enable information exchange between systems and their actors, using effectively and efficiently-usable way, there are two essential issues which need to be addressed: i) content structure - which addresses with the description of the content's inherent elements and also the relationship between these elements; and ii) communication interface - which addresses the interface through which each specific content will be exposed. There are a lot of eMaintenance ontologies which may help to solve these issues.

The aim of this paper is to: i) explore the main ontologies related to eMaintenance solutions and to inspect how Data Quality (DQ) aspect is considered within these ontologies; ii) propose a process called "Maintenance Data Production" and to find the relation between ontologies and their role during data production stages.

Keywords

eMaintenance, Ontology, Standards, Data Quality (DQ), ICT, Interoperability, Data Production.

1. INTRODUCTION

Interoperability is one of the most important aspects inside companies and business enterprises. It can be defined as the ability of applications and systems to share information and exchange services with each other based on standards and to cooperate in processes using the information and services [1]. IEEE has defined interoperability as the ability of two or more systems or components to exchange information and to use that information that has been exchanged [2].

Interoperability has many objectives. One important objective is the vision of software components working smoothly together, without regard to details of any component's location, operating system, programming language, or network hardware and software [3]. One solution to achieve interoperability is using standards. As essential property for long-term data retention, they offer stability in the way information is represented, and this retention issue is increasingly recognized as a costly and critical problem for industries with long product life cycles, such as aerospace [4].

As a result to the development in Information and Communication Technology (ICT), standards are required to

ensure performance, conformity, and safety of new products and processes. Standards are documented agreements containing technical guidelines to ensure that materials, products, processes, representations, and services are fit for their purpose. In case of data interoperability, standards are mainly described in the form of some data fixed formats are specified [5].

Standards, data exchange models and communication protocols are important aspects in order to achieve data interoperability between different systems in maintenance, operations and inside an organization hierarchy. When developing eMaintenance solutions as support to maintenance decision-making, integration architecture for data exchange between different data sources is important [6]. The design of integration architecture is highly depended on the mechanism that defines the structure of the data elements and also describes the relation between these elements, i.e. ontology. However, ontologies have a high impact on the integration architecture of eMaintenance solutions and affect its efficiency. The purpose of this report aims to investigate the state-of-the-art in ontologies related to maintenance.

Hence, to be interoperable, components and systems must correctly interpret words used as labels and data in an appropriate context. Today we are still far from achieving the essential levels of interoperability among manufacturing system components that provides significant improvements in manufacturing efficiency [7].

The purpose of this paper is to explore ontologies related to eMaintenance. The paper describes the focus domain for the investigated ontologies. Furthermore, the paper will study the data quality aspect within these ontologies and how that aspect is ensured during data exchange between systems.

2. MAINTENANCE DATA FLOW

Information and data quality as a competitive factor in Business Networking relates to the question of the extent to which information and data quality is decisive for the success of Business Networking efforts and therefore has a value in its own right [8].

To perform prognostic or diagnostic maintenance on a specific item, eMaintenance solutions require access to a number of different data sources, including maintenance data, product data, operation data, etc. As these sources of data often operate in a heterogeneous environment, integration between the systems is problematic [9]. As illustrated in Figure 1, different types of data are collected from heterogeneous sources, such as computer maintenance management systems (CMMS) and product data

management system (PDM). The data are processed and integrated through data fusion and transformed into eMaintenance information. Therefore, the quality of the data, or Data Quality (DQ), needs to be considered during the data collection process to the data fusion and integration process.

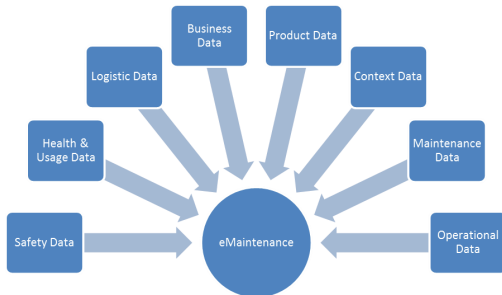


Figure 1, eMaintenance data access (Based on [9])

Since data often operate in heterogeneous environments, an important aspect for eMaintenance data is interconnectivity. All systems within the eMaintenance network must be able to interact as seamlessly as possible to exchange information in an efficient and usable way. As important aspects of DQ, Data accuracy, consistency and integrity should be assured within this network [10]. Accuracy of the data means that the recorded value is in conformity with the actual value [11]. Consistency can be defined as the representation of the data value should be the same in all cases [12]. Since issues in DQ in maintenance have direct impact to decision making process, all these aspects should be considered during the data flow process. This interconnectivity issues can be solved by applying eMaintenance ontologies.

Therefore, this paper will describe the different ontologies related to eMaintenance solutions and how each ontology can assure DQ.

3. EMaintenance AND DATA QUALITY

eMaintenance is a multidisciplinary domain based on maintenance and ICT, ensuring that the maintenance services are aligned with the needs and business objectives of both customers and suppliers during the whole product lifecycle [13]. eMaintenance can also be defined as a new concept that can be considered as a part of maintenance support. In this definition, eMaintenance provides information resources and information services which can be used to enable development and establishment of a proactive decision making process thought enhanced use of ICT [14]. eMaintenance solutions may include different services aimed for e.g. eMonitoring, eDiagnosis, and ePrognosis [15]. From a generic perspective, eMaintenance is maintenance managed and performed via computing [6].

Data Quality (DQ) can be defined as data that is fit for use by data consumers [16]. Information processing people considers that DQ is mainly focused on some attributes like accuracy, precision, and timeliness [17]. In maintenance decision-making, DQ is an important aspect to consider, since

without control of the DQ, there is no control of the accuracy of the output [18]. In a Computerized Maintenance Management System (CMMS), three important parties may affect DQ; *data producers*, *data custodians*, and *data consumers*. *Data producers* are people or systems that generate data. *Data custodians* are people who provide and manage computing resources for storing and processing data. Finally, *data consumers* are people or systems that use data. Therefore, data users are critical in defining data quality [19]. Thus, high information content (i.e., accurate, complete, and relevant information) leads to better product cost control and increased organizational efficiency (i.e., increased profit margin, increased decision making efficiency) [20]. As mentioned before, when developing eMaintenance solutions as support to maintenance decision-making, integration architecture for data exchange between different data sources is important [6]. Therefore, eMaintenance ontologies, like standards, have contributed in enhancing maintenance DQ. In this paper, the DQ aspect within these ontologies will be investigated in an attempt to identify the main contribution of these ontologies to the DQ.

4. DATA QUALITY ISSUES

In literature, DQ has been investigated extensively in prior information science research, where much of the discussion has been devoted to the underlying dimensions, such as accuracy, completeness, presentation, and objectivity where the main focus on describing DQ these dimensions or attributes [21].

Strong et al (1997) have observed 10 key DQ problems. Their observations were based on research that employed qualitative data collection and analysis techniques. In summary, these problems are [22]:

- 1) Multiple sources of the same information produce different values.
- 2) Information is produced using subjective judgments, leading to bias.
- 3) Systemic errors in information production lead to lost information.
- 4) Large volumes of stored information make it difficult to access information in a reasonable time
- 5) Distributed heterogeneous systems lead to inconsistent definitions, formats, and values.
- 6) Nonnumeric information is difficult to index
- 7) Automated content analysis across information collections is not yet available.
- 8) As information consumers' tasks and the organizational environment change, the information that is relevant and useful changes
- 9) Easy access to information may conflict with requirements for security, privacy, and confidentiality
- 10) Lack of sufficient computing resources limits access

In general, two types of information sources can be considered: subjective and objective. Subjective sources such as human observers, intelligence agents, newspaper reporters, experts and decision makers, supply observations, subjective beliefs, hypotheses, and opinions about what they see or learn. Quality of objective information sources such as sensors, models,

automated processes is free from biases inherent to human judgment and depends only on how well sensors are calibrated and how adequate models are [23].

In our previous study and from maintenance point of view, we have found that DQ issues can also be divided into two types: human related (subjective) and machine related (objective). Most of the identified DQ issues are subjective. For more information, see [24].

5. METHODOLOGY

The motivation to this study was the observed DQ issues from our previous research that has been mentioned before in section (4). In order to accomplish this study, we have divided the work into three parts. In part one, a literature study has been done to explore the eMaintenance ontologies in the different areas like condition based monitoring, product management, maintenance data management, data exchange etc. This study is done in section 4.1 eMaintenance ontologies. The second part of this study is to propose the maintenance data production process. In this process, data has been considered as a product. When we deal data as a product, DQ insurance will be easier to achieve. The last part of this study is to find relation between this process's stages and eMaintenance ontologies in order to produce high quality data. Part two and three is available in section 4.2

5.1 eMaintenance Ontologies

In eMaintenance solutions, the design of integration architecture mechanism defines the structure of the data elements and relation between these elements, i.e. ontology.

The term “manufacturing interoperability” refers to the ability to share technical and business information seamlessly throughout an extended manufacturing enterprise [4]. Published standards offer some stability by proposing information models for data representation, an essential property for long-term data exchange and archiving [25].

In this study, a literature review has been conducted in order to explore most of the available ontologies that are related to eMaintenance solutions. The studied ontologies are listed in Table 1 below. These ontologies are related to eMaintenance different solutions. The eMaintenance relation to these ontologies is stated in the right hand column of the table.

We can notice from Table 1 that eMaintenance ontologies have an important role to insure DQ in maintenance. They support maintenance DQ from data collection to the data visualization step. We can see the different eMaintenance areas related to these ontologies.

Table 1. Ontologies eMaintenance scope

Studied Ont.	Description	eMaintenance Scope
OPC UA	OPC is designed for Open Productivity and Connectivity in industrial automation and enterprise systems that support industry [26].	Software and Information interoperability
MIMOSA	MIMOSA provides metadata reference libraries and a series of information exchange standards using XML and SQL [27].	Measurement and condition based maintenance data transfer

PLCS	PLCS (Product Life Cycle Support) specifies an information model used for the exchange of assured product and support information throughout the entire product life cycle from concept to disposal [28].	Product Management and Maintenance, suitable to complex products and large companies
ISA-95	ISA-95 is the international standard for the integration of enterprise and control systems. ISA-95 consists of models and terminology to determine which information has to be exchanged between systems for maintenance and quality.	Maintenance Data Transfer and Management
XML	XML is a simple text-based format for representing structured information: documents, data, configuration, books, transactions, invoices, and much more [29].	Maintenance Data representation and exchange.
STEP	STEP is a family of standards defining a robust and time-tested methodology for describing product data throughout the life cycle of a product [30].	Product Life Cycle Management
CORBA	CORBA specifies interfaces that allow seamless interoperability among clients and servers under the object-oriented paradigm [31].	Services and Applications interoperability (objects)
OAGIS	OAGIS standard aims to achieve interoperability between disparate enterprise business systems by standardizing the architecture of the messages they exchange [32].	Data Interoperability between systems and databases
DPWS	DPWS is a common web services middleware and profile for devices, which defines two fundamental elements: the device and its hosted services [33].	Devices information exchange through web services
S1000D	It is an international specification for the procurement and production of technical publications. The S1000D provides ontology for the content of technical publications and also a content model, based on XML schema [34].	Technical Content Management
SOA	SOA represents a design framework for construction of information systems by combination of service. A service is a program unit which can be called by standardized procedures, and can independently execute assigned function [35].	Standardized data format for information and service interchange
SOAP	It defines a simple mechanism for expressing application semantics by providing a modular packaging model and encoding mechanisms for encoding data within modules [36].	Mechanism for exchanging structured and typed information between peers.
SCADA	It refers to a system that collects data from various sensors at factories, plants or in other remote locations and controls equipment over the SCADA networks [37].	Data collection and equipment control
ATA iSpec 2200	Is a global aviation industry standard for the content, structure, and electronic exchange of aircraft engineering and maintenance information. It consists of a suite of data specifications pertaining to maintenance requirements and procedures and aircraft configuration control [38].	Aviation industry
S4000M	S4000M is an International Specification for Scheduled Maintenance Analysis [38].	Maintenance Analysis
DAIS	Data Acquisition from Industrial Systems issued by the OMG is intended for online data transfer [2].	Data Acquisition

5.2 Maintenance Data Production

The Condition Based Monitoring (CBM) program to handle data and information in maintenance consists of three key steps; data acquisition, data processing and maintenance decision making [39]. This process in figure 2 includes data acquisition and processing before making a decision. Data acquisition is a process of collecting and storing useful data (information) from targeted physical assets for the purpose of CBM. Condition monitoring data are the measurements related to the health condition/state of the physical asset. It can be vibration data, acoustic data, oil analysis data, temperature, pressure, moisture, humidity, weather or environment data, etc. Various sensors, such as micro-sensors, ultrasonic sensors, acoustic emission sensors, etc., have been designed to collect different types of data. Wireless technologies, such as Bluetooth, have provided an alternative solution to cost-effective data communication. Maintenance information systems, such as computerized maintenance management systems (CMMS), enterprise resource planning systems, etc., have been developed for data storage and handling [39].

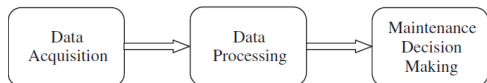


Figure 2, CBM program steps [39].

However, this process can be extended to add other fundamental aspects which need to be considered when establishing a maintenance information logistics solution. These new aspects are: content, time, communication and context [14].

Based on this process, we have proposed a process called “Maintenance Data Production” to refer to the process by which data is produced. This process describes all the actions that will occur to produce the data to the consumer as a final product. Maintenance Data Production (MDP) includes the following stages: data collection, data transition, compilation, analysis, visualization and contextualization.

These stages can be summarized as follows:

- 1) Data collection step: when obtaining the relevant data and managing its content. This data can be collected from different sources. These sources may be sensors, RFID tags, people etc.
- 2) Transition step: where the collected data need to be transferred without affecting its content. Data is transferred from source location to data management system.
- 3) Compilation step: to compile data from different sources in a way that insures its quality.
- 4) Analysis step: to analyse data and extract information and knowledge for decision making support.
- 5) Visualisation step: to visualise the information for the intended user or decision maker. The visualization could be statistical or reports.

- 6) Contextualization step: to put the visualized information into the needed context so it becomes meaningful and understandable in the right context.

During our study, we tried to find the relation between eMaintenance ontologies and the maintenance data production process. The results of this study are summarized in figure 3. Figure 3 shows that eMaintenance ontologies are required in every stage of the data production. In this figure, these ontologies are located in the stage that they may affect.

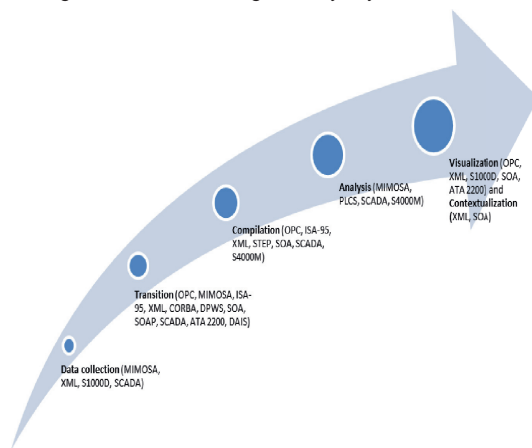


Figure 3, Maintenance Data Production Process

Hence, DQ management means that DQ should be insured during all these stages. From Figure 2, we can notice that eMaintenance ontologies have very important contribution in enhancing DQ.

6. RESULTS AND DISCUSSIONS

This study summarizes the available eMaintenance ontologies used to enhance maintenance DQ. When accomplishing the study, we can say that eMaintenance solutions are widely used and they have a lot of contributions in different industry and business aspects.

From this study, it was clear that eMaintenance ontologies contributions are represented mainly with DQ enhancement during the process of data production phases from data collection to data visualization and contextualization process. In addition, the studied ontologies were divided into the different areas it may affect.

The other contribution of this study is the suggested process that is called “Maintenance Data Production”. This process considers all the stages required to produce the data as a final product. An attempt to investigate which eMaintenance ontologies are used during each stage has been done.

When applying eMaintenance ontologies to the maintenance data production process, the wide range of available ontologies makes it difficult to decide which ontology that is best suited for each stage in the Data Production process. In this article, an

attempt to find the relation between the ontologies and the area that ontologies should be used in has been done. The results are shown in Table 1 above.

By applying ontologies to eMaintenance solutions according to the areas described in Table 1 and the stages in the Data Production Process, we can insure that we will get a high quality data and that leads directly to obtain an adequate and effective decision making.

7. CONCLUSIONS

Interoperability and data quality are crucial and indispensable for both maintenance and operations. They are also critical for achieving higher levels of organizational interoperability that are required for an effective decision making. In order to increase the economic benefit and enhance decision making, the use of eMaintenance tools should be adapted in more effective use. Standards, models and services discussed in this paper are offering a sustainable support for this objective. From our study, we can conclude the following:

- 1) DQ issues in maintenance have direct impact to the decision making that leads to wrong decisions regarding the maintenance process.
- 2) All discussed standards and services provide important support to the interactions necessary to construct unified maintenance and operations and enhance integration among systems of different origins to achieve integrity that helps in ensuring data quality.
- 3) The proposed Maintenance data production process provides a clear view to the steps that should be followed in order to produce a high quality data.
- 4) Applying eMaintenance ontologies in each step of the maintenance data production process helps to insure DQ.
- 5) Different challenges still available need to be discussed more effectively in order to help in developing adapting the available eMaintenance solutions. Some of them are represented by the high cost and the time that prevent some companies from utilizing these ontologies. Another challenge could be the lack or the ineffective use of these tools in many enterprises and organizations. Finally, the wide range of available ontologies makes it difficult to decide which ontology that is best suited for each stage in the Data Production Process.

8. REFERENCES

- [1] E. Murphy, "OPC enables interoperability," *Hydrocarbon Process.*, vol. 86, no. 3, pp. 51-52 2007.
- [2] OMG. *Data Acquisition from Industrial Systems* [Online]. available: www.omg.org.
- [3] J. Siegel, "OMG overview: CORBA and the OMA in enterprise computing," *Commun ACM*, vol. 41, no. 10, pp. 37-43 1998.
- [4] S.R. Ray and A. Jones, "Manufacturing interoperability," *J.Intell.Manuf.*, vol. 17, no. 6, pp. 681-688 2006.
- [5] R.H. Allen and R.D. Sriram, "The role of standards in innovation," *Technological Forecasting and Social Change*, vol. 64, no. 2, pp. 171-181 2000.
- [6] M. Kajko-Mattsson, R. Karim and A. Mirijamdotter, "Fundamentals of the eMaintenance Concept," in *1st international workshop and congress on eMaintenance*, 2010, pp. 22-24.
- [7] J. Klobas, "Knowledge management in small enterprises," *The Electronic Library*, vol. 18, no. 6, pp. 420-432 2000.
- [8] B. Otto, Y.W. Lee and I. Caballero, "Information and data quality in business networking: A key concept for enterprises in its early stages of development," *Electronic Markets*, vol. 21, no. 2, pp. 83-97 2011.
- [9] K. Wandt, R. Karim and D. Galar, "Context adapted prognostics and diagnostics," in *International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, London, UK, 2012.
- [10] R. Blake and P. Mangiameli, "The effects and interactions of data quality and problem complexity on classification," *Journal of Data and Information Quality (JDIQ)*, vol. 2, no. 2, pp. 8 2011.
- [11] D.P. Ballou and G.K. Tayi, "Enhancing data quality in data warehouse environments," *Commun ACM*, vol. 42, no. 1, pp. 73-78 1999.
- [12] D. Ballou, S. Madnick and R. Wang, "Special section: Assuring information quality," *J.Manage.Inf.Syst.*, vol. 20, no. 3, pp. 9-11 2003.
- [13] M. Kajko-Mattsson, R. Karim and A. MIRJAMDOTTER, "Essential components of e-maintenance," *International Journal of Performability Engineering*, vol. 7, no. 6, pp. 555 2011.
- [14] R. Karim, *A service-oriented approach to e-maintenance of complex technical systems*, Division of Operation and Maintenance Engineering, Luleå University of Technology, 2008.
- [15] U. Kumar, "Special issue on eMaintenance solutions and technologies guest editorial," *International Journal of Systems Assurance Engineering and Management*, vol. 1, no. 3, pp. 187-188 2010.
- [16] D.M. Strong, Y.W. Lee and R.Y. Wang, "Data quality in context," *Commun ACM*, vol. 40, no. 5, pp. 103-110 1997.

- [17] A.V. Levitin, "Data as a resource: Properties, implications, and prescriptions," *Sloan Manage.Rev.*, vol. 40, no. 1, pp. 89-101 1998.
- [18] R. Karim, M. Kajko-Mattsson and P. Söderholm, "Exploiting SOA within emaintenance," in *Proceedings of the 2nd international workshop on Systems development in SOA environments*, 2008, pp. 75-80.
- [19] R.L. Leitheiser, "Data quality in health care data warehouse environments," in *System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on*, 2001, pp. 10 pp.
- [20] N. Gorla, T.M. Somers and B. Wong, "Organizational impact of system quality, information quality, and service quality," *The Journal of Strategic Information Systems*, vol. 19, no. 3, pp. 207-228 2010.
- [21] O. Arazy and R. Kopak, "On the measurability of information quality," *J.Am.Soc.Inf.Sci.Technol.*, vol. 62, no. 1, pp. 89-99 2011.
- [22] D.M. Strong, Y.W. Lee and R.Y. Wang, "10 potholes in the road to information quality," *Computer*, vol. 30, no. 8, pp. 38-46 1997.
- [23] G.L. Rogova and E. Bosse, "Information quality in information fusion," in *Information Fusion (FUSION), 2010 13th Conference on*, 2010, pp. 1-8.
- [24] M. Aljumaili, V. Rauhala, P. Tretten and R. Karim, "Data quality in eMaintenance : a call for research," in *International Conference on Maintenance Performance Measurement & Management : MPM*, Luleå, Sweden, 2011, pp. 69.
- [25] A.F. Cutting-Decelle, L. Pouchard, B. Das, R. Young and J. Michel, "Utilising standards based approaches to information sharing and interoperability in manufacturing decision support," in *International Conference FAIM04, Toronto, Canada*, 2004.
- [26] OPC. [Online]. available: www.OPCfoundation.org.
- [27] MIMOSA. [Online]. available: www.mimosa.org.
- [28] Eurostep. [Online]. available: www.plcs-resources.org.
- [29] W3C. [Online]. available: www.w3.org.
- [30] R.S. Peak, J. Lubell, V. Srinivasan and S.C. Waterbury, "STEP, XML, and UML: Complementary technologies," *Journal of Computing and Information Science in Engineering*, vol. 4, no. 4, pp. 379-390 2004.
- [31] I. Pyarali and D.C. Schmidt, "An overview of the CORBA portable object adapter," *StandardView*, vol. 6, no. 1, pp. 30-43 1998.
- [32] X. Fiorentini and S. Rachuri, "STEP-OAGIS harmonization joint working group," 2009.
- [33] G. Cândido, F. Jammes, J.B. de Oliveira and A.W. Colombo, "SOA at device level in the industrial domain: Assessment of OPC UA and DPWS specifications," in *Industrial Informatics (INDIN), 2010 8th IEEE International Conference on*, 2010, pp. 598-603.
- [34] R. Karim, O. Candell and P. Söderholm, "E-maintenance and information logistics: Aspects of content format," *Journal of Quality in Maintenance Engineering*, vol. 15, no. 3, pp. 308-324 2009.
- [35] N. Komoda, "Service oriented architecture (SOA) in industrial systems," in *Industrial Informatics, 2006 IEEE International Conference on*, 2006, pp. 1-5.
- [36] D. Box, D. Ehnebuske, G. Kakivaya, A. Layman, N. Mendelsohn, H.F. Nielsen, S. Thatte and D. Winer, *Simple object access protocol (SOAP) 1.1* 2000.
- [37] D. Choi, H. Kim, D. Won and S. Kim, "Advanced key-management architecture for secure SCADA communications," *Power Delivery, IEEE Transactions on*, vol. 24, no. 3, pp. 1154-1163 2009.
- [38] ATA. [Online]. available: www.airlines.org.
- [39] A.K.S. Jardine, D. Lin and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical systems and signal processing*, vol. 20, no. 7, pp. 1483-1510 2006.

Information Security in eMaintenance – A study of SCADA Security

Yamur K. Al-Douri

Department of Civil, Environment
and Natural Resources Engineering,
Luleå University of Technology,
P.O. BOX 97187, Luleå,
S-972 31 Sweden

yamur.aldouri@ltu.se

Mustafa Aljumaili

Department of Civil, Environment
and Natural Resources Engineering,
Luleå University of Technology,
P.O. BOX 97187, Luleå,
S-972 31 Sweden

mustafa.aljumaili@ltu.se

Ramin Karim

Department of Civil, Environment
and Natural Resources Engineering,
Luleå University of Technology,
P.O. BOX 97187,
Luleå, S-972 31 Sweden

ramin.karim@ltu.se

ABSTRACT

eMaintenance solutions are spreading increasingly due to the continuous evolution in the different Information and Communication Technology (ICT) tools. In general, most of the available eMaintenance solutions are depending on Internet infrastructure what makes them vulnerable to all security threats that affect the Internet. One of the important eMaintenance solutions is Supervisory Control and Data Acquisition (SCADA) system as it has been used in most of the industrial processes. SCADA systems were designed without security considerations as they were mainly installed into isolated networks. Nowadays, SCADA systems are mainly connected to Internet and other networks. Therefore, SCADA systems have been exposed to wide range of network security threats. Hence, SCADA security has become an important aspect that needs to be investigated. In this paper, a study of SCADA security issues will be done. The main contribution of this paper is to address SCADA security issues and challenges related to eMaintenance.

Keywords

SCADA, eMaintenance, Intrusion Detection System, Artificial Immune System, Firewall, Cloud Computing.

1. INTRODUCTION

Nowadays, variant industrial systems have to cooperate together over hundreds of devices and various software systems in order to achieve their goals and to develop industry [1]. SCADA is Supervisory Control and Data Acquisition system designed to use the currently available long-distance telecommunications technologies. SCADA is used to monitor and remotely control critical industrial processes [2].

Within SCADA, a communications protocol is used to deliver successfully a message from one computer/device to another using the serial communication channel. These devices like PLC (Power Line Communication) are an electronic device used for industrial processes automation. Protocols like any other communication services have some challenges and security is one of the most important challenges [3].

The fast development in ICT (Information and Communication Technology) is a vast broad subject added velocity to everything that is done within industry. Worthwhile during last years, we need to mention that these developments have been incorporated

into the development of maintenance activities [4]. Along with this development, eMaintenance has been widely used as an improvement of productivity and optimization of costs through utilization of Web services [4] [5] [6].

eMaintenance can be considered as a maintenance strategy where different tasks are managed electronically by the use of real-time item data, such as mobile devices, remote wireless sensing, condition monitoring, knowledge engineering, telecommunications and internet technologies [4] [5].

SCADA network has increasingly supported eMaintenance system in order to improve maintenance data collection process by improving the network quality. However, SCADA system has a lot of security issues. Some of these vulnerabilities are external user input, such as browsers or mail software, may allow external parties to take control of a device [7]. Therefore, security of SCADA networks has become a prime concern because an attack against a SCADA might affect the system safety and availability [8]. The goal of this paper is to investigate SCADA security issues and challenges focusing on those are related to eMaintenance.

This paper has nine sections. The first one is introduction; 2. SCADA architecture; 3. methodology described the study surveys, 3. SCADA Attacks general study of the attacks related to industry, 4. SCADA Models Validity to gather experimental model testing, 5. Challenges of SCADA security and variant existed solutions of the security problems in section 6. Finally, 7. Conclusion and Future work. Overview to most researches mention in security mechanism.

2. SCADA ARCHITECTURE

In the 1960s, SCADA system used to available for long distance telecommunications technologies of the time either the Telephone Company ("Ma Bell") technology or radio technology [2].

SCADA is Supervisory control and data acquisition are used as control critical industrial processes for a monitoring, such as electric power transmission, gas pipelines, and potable water distribution/delivery [2].

There are two types of SCADA: *SCADA hardware* which consists of a number of Remote Terminal Units (RTUs) collecting field data and sending these data back to a master station, by using a communication technique; and *SCADA software* where companies develop software to communicate to their hardware. In

addition, there are different features of SCADA software like User interface, Graphics displays, Alarms, and RTU (and PLC) interfaces, [9]. Therefore, all these components are cooperating together to produce SCADA network.

SCADA system has different components cooperate together to produce a SCADA network. The primary mission of a communications protocol is to successfully deliver a message from one computer/device to another. These devices like RTUs are able to generate control outputs. This output control signals are initiated or adjusted on receipt of a command from the central computer, usually initiated by a human operator. Also, PLC (Power Line Communication) is an electronic device used for industrial processes automation. The PLCs were distributed and the systems became more intelligent and smaller in size comparing with other Monitor and control devices in the plant grew to improve the products reliability [2] [9] [10] [11]. While HMI is human operator used to control operations. It has been a necessary requirement to monitor and control multiple RTUs remotely, and other control devices. Usually, PLCs are enabling technicians to configure HMIs themselves without using any tailor-made standard software packages [10].

There are about 1500-2000 SCADA protocols developed by deferent companies [8]. Currently, the list of IP-compatible RTU protocols are fairly short: DNP3.0, ICCP, UCA2.0, and Modbus [12]. Security is one of the most important issues for protocols [3]. Different researches tried to develop encryption algorithms to encrypt protocol messages.

The formal of SCADA systems were designed without security because mainly installed into isolated networks [8]. There are different organizations are responsible to improve different standards for SCADA security. The Instrumentation Systems and Automation (ISA) society has security SCADA industry guidelines. This organization defines and published the standards of security consist with the industry. Like OPC is an organization used open standards for industry. Also, The National Institute for Standards and Technology (NIST) for control centre in SCADA network using Common Criteria.

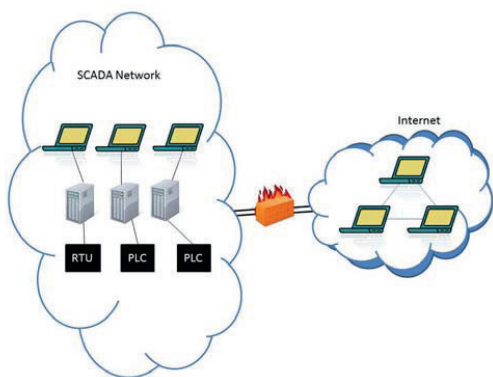


Fig. 2-1: SCADA Network Architecture

Figure 2-1 shows the SCADA network architecture. The main advantage of SCADA is its suitability to a wide range of

applications starting from nuclear power plants to controlling of room temperature due to its high compatibility and reliability. Unfortunately, SCADA also has some issues. The main problem of SCADA applications is that they come with lack of standardized protocols due the legacy of early low bandwidth [10].

3. METHODOLOGY

The paper structure is gathering most literatures in each SCADA security existed solutions mentioned in this paper. There is a different number of interesting literature depends on the mechanism. Rather mechanism does not have wide range of literatures based on the new used technologies. We found 126 papers with different aspects. They have a suggestion model and deep interesting look to the mechanism. We will discuss all the percentages for each mechanism. The figure below (3-1) describes the interesting literature in each technology.

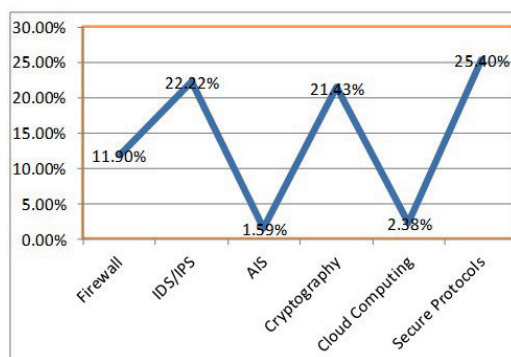


Fig. 3-1: SCADA Security mechanisms survey

The figure above portrays, there are a slightly literatures related to SCADA network with Cloud computing and AIS. So we should focus on this section as a future work. In other hand, the literatures focus on securing protocols governs the transmission in SCADA channels using different encryption algorithms. That result also in cryptography has near accuracy because securing protocols based on encryption algorithms. While, IDS/IPS has a good accuracy due the focus of the literatures to secure the contents through the transmission based on the rules produced.

4. SCADA ATTACKS

Generally, Risks can be defined as the probability that a threat will exploit system weakness. While vulnerabilities are external user input as a third party, like browsers or mail software, may allow external parties to take control of a device. Then, it will create an effect damaging to the system. Threat is a possible source of danger caused by human or person to any system to exploit vulnerable [7].

We should define Cyber-attack in order to adopt the solution for intrusions in the SCADA network. Cyber-attack is a computer-to-computer attack. The security of SCADA network objectives must be carefully looked at which are maybe in contradiction to

have still the fast action of an authorized person. We should ensure with different objectives achieved that [13].

There are three main threats attack SCADA network [14]:

- 1) **Response Injection:** this attack injects into a control system a false response.
- 2) **Command injection:** the attacker injects false control commands.
- 3) **Denial of Service (DOS):** the attacker attempt to break and cut the communication link between the remote terminal device and master or human machine interface.

Industry risks like SCADA attack is StuxNet (2010): this is a complex design threat. It specially designed for industrial systems to include with PLC rootkit. It has a capability to reprogram the PLCs and hide its changes [13].

There are eight security strategies define all the security requirements for the system [15] [16]. These objectives are Confidentiality, Integrity, Availability, Authorization (access control), Authentication, Non-reputability, Auditability, and Third party protection [16].

Security mechanisms used for SCADA have different ways. That might include security policies, firewalls and intrusion detection systems/ prevention system, the use of encryption, multi-layered authentication, Cryptography and the use of virtual private networks (VPN). These mechanisms used to achieve the previous objectives [17].

5. SCADA MODELS VALIDITY

Modern SCADA networks used familiar protocols like TCP. These types of networks we can expertize the validity of our systems through Knowledge Discovery in Database (KDD Cup data) provided by the Defence Advanced Research Projects Agency DARPA and the MIT Lincoln Labs. Originally, the data consisted of nine weeks of raw Transmission Control Protocol (TCP) dump data from the network. The dataset contains normal and attacks labels. KDD'99 dataset based on DARPA 98. The KDD'99 data set contains huge number of redundant records. To solve these problems, all the repeated records in the entire KDD will be removed. This led to the establishment of a new data set NSL-KDD that consists of selected records of the complete KDD data set and does not suffer from any of mentioned deficiency. NSL-KDD data set is employed to avoid redundant records which may cause learning algorithm to be biased [18].

It is very much necessary to move away from static datasets toward more dynamically generated datasets which not only reflect intrusions and the traffic compositions of that time, but are also modifiable, and reproducible. Systematic approach for this required generate datasets is introduced to address this need. This dataset is based on the concept of profiles which contain abstract distribution models for applications and detailed descriptions of intrusions, or protocols. Real traces are analysed to create agents profiles that generate real traffic for HTTP, SMTP, IMAP, and FTP. The profiles were generated employed in an experiment to generate the desirable dataset in a testbed environment. Variant scenarios of multi-stage attacks were subsequently carried out to supply the anomalous portion of the dataset. This dataset is better than other datasets such as CAIDA (2011) [19].

Also, there is Industrial Networking Security laboratory. This laboratory was created to a cooperative research activity with a prominent power company. The typical power architecture plant is complex and consists of pipes, valves, sensors, pumps, etc. The SCADA server is directly connected to other equipment that used to control real power plants process networks. Some surveys used Modbus over TCP protocol and PLCs of the ABB AC800 family [14]. AC800 is safe and efficient application evolution. To reuse your installed I/O, we use superior methods for controller application conversion [20].

The side of SCADA simulator, there are different software used to simulate this network. Like Consipio Software LLC is a Texas Corporation specialized. It is a tool that is used to facilitate operations of SCADA System like energy companies and utilities.

Finally, SCADA system should diagnosis all of these vulnerabilities and test the values of security mechanisms.

6. SCADA SECURITY CHALLENGES

This section provides an overview of the main research challenges in the SCADA security. The challenges are [21]:

- Access controls improvement to the SCADA network. That improve will prevent any attacker to enter to SCADA network.
- Security monitoring tools then improve that inside the SCADA network. These tools will help to trace the attacker and prevent any change in the data.
- Security management improvement of the SCADA network.
- Authentication providing, fine granular file system protection and memory isolation between processes.
- Newly philosophy of implemented systems. SCADA network tend to have very long lifetimes.

7. SCADA SECURITY SOLUTIONS

There are many existed solutions to secure the SCADA network. These solutions help to avoid the challenges. This section will mention to these solutions to achieve Security SCADA network.

7.1 Firewall

Generally, single access point runs on a device as a Firewall consists of several different filters. That used to block transmission of specific classes of traffic [22]. That also helps to maintain the system from any future failure.

Firewall placed between an organization and the outside world to protect a local system or network of systems [22] [23]. As we notice in Fig. 1-1 the Firewall prevents foreign attacks to let inside SCADA network.

Firewall classified into three main categories:

- 1) **Packet filtering:** it is gateway security, need just router to connect to internet without and another charges because it is easy and cheap. As an example ISA server dropping packets based on their source, destination, both, or port [23].
- 2) **Circuit gate ways:** it can be a specialized function performed by an application-level gateway for specific applications. It is stand-alone system [23].

- 3) **Proxy Server (Application gateway)**: different clients connect to the server in secure way and without affecting to anything using a TCP/IP application. It is done by make intelligent decisions about information through understands application data and can therefore [24].

The firewalls have their limitations: it cannot protect internal systems may have dial-out capability to connect to an ISP [22] [23].

There are a complex firewall detects attacks to an internal controller of the SCADA system. There is a new approach based on Master-Slave communication to filtering and analysis malicious packets based on Critical State-based Analysis for analysing Modbus and DNP3 traffic. This firewall based on classical signature-based approach. Firewall should cover the gaps through: 1) knowing the actual state of that system, 2) control the architecture of the system, 3) the flowing operative of the SCADA commands between master and slaves, and 4) the set of unwanted (critical) states. That will help the firewall his steps to manage and prevent the attackers to login into SCADA network [14].

Recently, (NISCC) the UK government's National Infrastructure Security Coordination Centre released different guidelines for efficient side use of firewalls in SCADA networks. The suggested firewall can detect variant protocols, Denial of Services (DOS) attack and other types of attacks [8] [26].

Another Firewall improved Multi-Threat protection is FortiGate-5000 Series. That will block threats and network attacks like Trojans, worms, phishing schemes, intrusion attempts, denial of service (DoS), and different viruses [26].

Additionally to improve SCADA Security, it should add some technologies to the network like Virtual Private Networks (VPN), Virtual LANs (VLAN), and Demilitarized Zones (DMZ). These technologies beside Firewall will improve the security of SCADA system [27].

7.2 Intrusion Detection & Prevention System (IDS/IPS)

Intrusion Detection Systems (IDSs) works like a safeguard, it detects a violation of its configuration (similar to an opened or broken window) and activates an alarm [22]. IDS used to monitoring the system and alarm any unknown packet to enter the network organization [22] [28]. Developing IDS is more difficult than developing firewalls for SCADA networks. The rules developed to recognize attacks require a good knowledge of the vulnerabilities in the protocols of SCADA network [8]. The IDSs are mostly installed: between network and extranet, before the firewall to identify the attacks on the server [22].

There are three main types of IDSs, depending on how they monitoring activity:

- 1) **Host-Based Intrusion Detection System (HIDS)**: to check only activity on a specific host, system audit, and checks event logs from different operating systems and applications on a specific host [22].
- 2) **Network-Based IDS (NIDS)**: analyse all the packets on a particular network looking for attacks. It focuses on network

traffic flowed along the connection that interconnect the systems [28].

- 3) **Hybrid of HIDS and NIDS**: combination between previous two types to increase usability by take advantage of each one, also decrease the disadvantages [28].

In addition, there are three techniques (methods) used to construct IDS:

- 1) **Signature-Based IDS (Misuse Detection)**: using pattern matching for each packet to detect and find known attack patterns [28].
- 2) **Normal Behavior Patterns (Anomaly Detection)**: this method gives better result compared with the signature method. That because it depends on the differences of the packets which are present in the parts of the protocol header [22].

The previous two techniques used with SCADA network help to improve the SCADA Security [29]. Also, design a security pattern tool that will secure SCADA system after analyze the potential attacks against it. This tool like Role-Based Access Control (RBAC) pattern combined with the Authenticator Pattern and the Logger Pattern. That will build a new Firewall based on IDS to get a new direction of SCADA security [30].

Therefore, multiple overlapping layers of security are used to protect critical assets as data logger. That will responsible to capture, time stamps, cryptographically signs, encrypts, and store network traffic. The data logger was validated in the Mississippi State University SCADA Security Laboratory. Then, this data logger support use of signature based technique. Human Machine Interface (HMI) used to control and monitor the system [31].

7.3 Artificial Immune System (AIS)

Artificial immune system is a computational intelligence used to recognize cells in the body, and divide these cells into two groups. One group is self-named antibody, and the other is non-self-named antigen, which is eliminated when it is classified [32] [33]. Antigen could be defined as foreign invaders find a way a considered system [33]. While Antibodies are parts of the system used to detect and eliminate antigens by catch them [32] [33].

Particularly, the detection and prevention antigens which affect Industrial Process Control (IPC) systems such as SCADA. The SCADA Network equipment like latest PLC technology monitor process behaviour and all software updates. PLCs give us the ability to access our control system to handle such tasks as monitoring via a website to determine the condition of a machine or check other statistics like different messages. In order to detect malware of the PLCs through include a secure Virtual Machine (VM) or a dedicated host to the SCADA Network. This virtual machine detects any unauthorized or abnormal behaviour. VM will monitor the Main System and RTUs/PLCs for abnormal activities. In other hand, to increase protection, a virtual private network (VPN) should be used in the SCADA network. That will help to encrypt the transmitted data in safe side when traveling over a public network (as the Internet) [34].

7.4 Cryptography and Cryptanalysis

Different sufficient diverse measurement is useful to reduce the risks of a successful attack from outside the system. Encryption

and cryptanalysis technology used to prevent a third party from prying into the contents of messages being passing electronically through the processes. The processing of the message before sending is encryption to be understood in the other side receiver [35]. Typically, SCADA protocols do not support any sort of cryptography. Although, there are challenges to implement cryptography techniques in SCADA network due to the limitation of computational capabilities of SCADA devices [8].

Generally, there are two types of encryption [36]:

- 1) **Symmetric encryption (private-key encryption)**, the sender and the receiver use the same secret key. The secret key will be responsible to controls the encryption process.
- 2) **Asymmetric encryption (public-key encryption)**, this technique uses two different keys. The first one used by sender to encrypt the message, and the other one used to decrypt the message in the receiver [35].

As we mention in Introduction section. Protocol is the manager of the messages sent by the SCADA communication networks. We will mention to the new generation of protocols through encrypt the message transmitted through SCADA network.

Unfortunately, SCADA protocols for were designed without security in mind. Currently, there is a reasonable solution to secure Modbus/TCP protocol. It is one of the Modbus family protocols. Normally, via a firewall or router access control lists (ACL) filter TCP port 502. This type of rule only permits or denies Modbus/TCP traffic from a given source to a given destination [37]. In other hand, there are number of paths help to develop a secure Modbus. Integrate security mechanisms into actual Modbus protocol through using different encryption algorithms. Finally, implement the above security protocol on the system [15].

The importance of Modbus protocol leads the researchers to use Predictive YASIR encryption to evaluate Modbus/ASCII protocol. That has a quite encryption technique and some latency by using this encryption [38].

Some surveys discussed the security protocols. To achieve security for protocols should build blocks of cryptographic algorithms or hash function keys. That implemented through add key management within SCADA security. Also, they classify key management depends on communication modes. As a Master-Slave communication, this type of communication based on the DNP3.0 or ISO/IEC 11770. Two sides have a pre-shared symmetric key which is used to encrypt and decrypt the transmission message depends on the type of the protocol. Peer to peer is other type of communication has different algorithms used for the encryption. TTP algorithm distributes the shared secret key between A and B as a third party. When the key translation centre (KTC) of the TTP generated shared key between A and B (session key). Then, the session key is encrypted by the pre-shared key [39].

DNP3 protocol is widely used in process industry. Securing that by DNP3Sec as a framework through built a BITW (Bump-in-the-wire). This BITW is a prototype which exchanges keys and uses those keys to encrypt the data flow in the network. DNP3Sec protocol used to protect the data without increasing the message size or deleting any data from the packet header. The BITW encapsulates DNP packet within DNP3Sec and adds

Authentication Data for integrity check. This type of protocol will increase the security of the SCADA network without effect on the SCADA system [40].

One of the most important challenges of SCADA network is to produce a secure protocol used with low speed serial communication protocol. AES-GCM encryption algorithm is used to implement I2CSec protocol with peer-to-peer communication type [41].

As an examples of using cryptography in SCADA network is American Gas Association (AGA). That will increase the ability of system's to detect malicious cyber traffic with objective to minimize cyber vulnerabilities. AGA 12 Cryptography developed a suite of open standards to protect the data transmitted by SCADA systems, which will authenticate the messages on SCADA systems, ensure data integrity, and maintaining the performance requirements of the SCADA link over serial SCADA network [42].

There are wide implementations of SCADA protocols with security using different encryption algorithms to increase the data security and decrease the latency of sending packets. We mentioned popular protocols researchers focused on.

7.5 Security Cloud Computing

New technology is Cloud computing, and that will re-change information technology (IT) trend. Through using Cloud technology, users are working over the internet through virtualizing variant of devices, PCs, laptops, smartphones, and PDAs to access programs, storage, and application platforms, via services offered by cloud computing providers [43] [44]. Adding Cloud computing for SCADA will improve the reliability via multiple input connections and will be easy to update and manage remotely [45].

There are three types of cloud computing [44]:

- 1) **Public cloud:** the resources are dynamically provisioned over the Internet through services or applications.
- 2) **Private cloud:** for exclusive use of one client, they have special providing full control over data, security, and quality of service.

Cloud computer security can embed Firewall, Demilitarized Zones, Network Segmentation, Intrusion Detection and Prevention Systems, and other tools [44].

Through the combination between Cloud Computing technology and SCADA that will improve the availability and easy scalability by include cost savings, energy saving, less space requires, management [43].

The corporation between SCADA and Cloud computing appears in Figure 7-1.

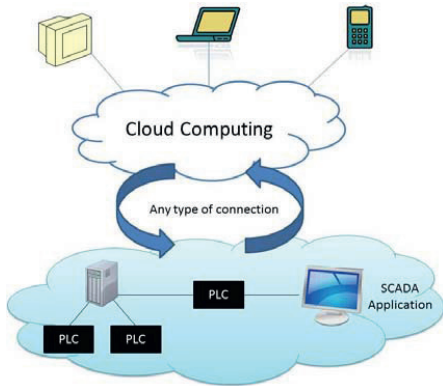


Fig. 7-1: Cloud SCADA system

Provide SCADA with Cloud services for their business that take to use their own IT infrastructure [45]:

- 1) Directly connected SCADA application to the control network through Public Cloud. Then, the information delivers to the specific place in the cloud where it can be stored or processed.
- 2) SCADA application running entirely in the Private Cloud, then remotely connected to the control network via WAN.

There are many challenges govern the cloud computing [43]. The most important challenge for our research is *Security and Privacy*: there is different interesting papers concern about security cloud computing. The dangers from the attacks still stored distant from the firewall [43].

Cloud SCADA security takes different fashions of protocols encryption to keep the transmission safe. Also, there is different mechanism used to keep transmission secure.

8. CONCLUSION AND FUTURE WORK

Generally through this study, eMaintenance is improvement of the functionality through utilization new technologies. Normally, it is expand of maintenance strategy. eMaintenance support resources, services and management necessary to able to monitor plant floor assets through link the production and maintenance operation systems. That will fix future problems (Risks) by predict before the failure of the system. eMaintenance has different aspects used to improve the decision of SCADA system. SCADA is used to monitor and remotely control critical industrial processes through using possible devices like RTU, PLC, PAC, and others.

The main challenge of SCADA network is vulnerabilities are external user input prevent the system functionality based Cyber-attack. Securing SCADA protocol systems are important solution to avoid and predict any system failure using cryptography technique. Additionally, there is different security tools used to ensure of the data safety and keep monitoring systems safe through variant mechanisms within SCADA. That might include security Policies, Firewalls and Intrusion Detection Systems/Prevention System, Artificial Immune system, and Cloud

Computing. These tools beside organizations standards will produce different rules and policies to manage and control the relationship between the people and other ICS (Industrial Control System) components.

Future vision is suppose a new model recognizes attacks based on the new technologies like cloud computing with AIS (Artificial Immune System). The cloud computing will utilize the cost of the SCADA system and in other hand achieve an artificial rules.

9. REFERENCES

- [1] Zhang, W., Halang, W., Diedrich, C. (2003), "An agent-based platform for service integration in e-maintenance", IEEE International Conference on Industrial Technology, Vol. 1, pp. 426-433.
- [2] Choi, D., Lee, S., Won, D., Kim, S. (2008), "Efficient Secure Group Communications for SCADA", IEEE Transactions on Power Delivery, Vol. 25, Issue. 2, pp. 714-722.
- [3] Tang, W. and Sui, A. (2008), "Secure protocol lifecycle and its application in power industry", Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT), pp. 592-597.
- [4] Karim, R. (2008), *A Service-Oriented Approach to eMaintenance of Complex Technical Systems*, Luleå University of Technology, Sweden.
- [5] Mullera, A., Marquez, A. C., Iunga, B. (2008), "on the concept of e-maintenance: Review and current research", *Reliability Engineering & System Safety Journal*, pp. 1165-1187.
- [6] Kumar, U., Parida, A., Karim, R. (2010), "Special issue on eMaintenance solutions and technologies: guest editorial", *International Journal of Systems Assurance Engineering and Management*, Vol. 1, pp. 187-188.
- [7] Gollmann, D. (2011), *Computer Security*, 3rd edition, John Wiley and Sons Ltd.
- [8] Ijure, V. M., Laughter, S. A., Williams, R. D. (2006), "Security issues in SCADA networks", *Computers & Security journal*, Vol. 25, Issue 7, pp. 498-506.
- [9] Bailey, D. and Wright, E. (2003), *Practical SCADA for Industry*, 1st edition, Newnes Publishers.
- [10] Chowdhury, S., Crossley, P., Chowdhury, S. P., (2009), *Microgrids and Active Distribution Networks*, Institution of Engineering and Technology.
- [11] Daniel, K. (2010), *Programmable Automation Technologies: An Introduction to CNC, Robotics and PLCs*, Industrial Press; 1 edition.
- [12] Shaw, W. (2006), *Cybersecurity for SCADA Systems*, PennWell Publishers.
- [13] Daniela, T. (2011), "Communication security in SCADA pipeline monitoring systems", 10th Roedunet International Conference (RoEduNet), pp. 1-5.
- [14] Fovino, I.N., Coletta, A., Carcano, A., Masera, M. (2012), "Critical State-Based Filtering System for Securing SCADA Network Protocols", IEEE Transactions on Industrial Electronics, Vol. 59, Issue. 10, pp. 3943-3950.

- [15] Byres, J. E., Franz, M., Miller, D. (2004), "The Use of Attack Trees in Assessing Vulnerabilities in SCADA Systems", International Infrastructure Survivability Workshop (IISW'04), IEEE.
- [16] Choi, D., Kim, H., Won, D., Kim, S. (2009), "Advanced Key-Management Architecture for Secure SCADA Communications", IEEE Transactions on Power Delivery, Vol. 24, Issue 3, pp.1154-1163.
- [17] Nicholson, A., Webber, S., Dyer, S., Patel, T., Janicke, H. (2012), "SCADA security in the light of Cyber-Warfare", *Computers & Security Journal*, Vol. 31, Issue 4, pp. 418-436.
- [18] Tavallae, M., Bagheri, E., Lu, W., A., A. G. (2009), "A detailed analysis of the KDD CUP 99 data set", CISDA 09, IEEE Symposium on Computational Intelligence for Security and Defense Applications.
- [19] Shiravi, A., Shiravi, H., Tavallae, M., Ghorbani, A. A. (2012), "Toward developing a systematic approach to generate benchmark datasets for intrusion detection", *Computers & Security Journal*, Vol. 31, Issue 3, pp. 357-374.
- [20] ABB (2010), *Migrate your MP200/AC450 controller to the AC800 controller with minimum impact on production*, Available at: [http://www05.abb.com/global/scot/scot227.nsf/veritydisplay/3aa1cbc01457c20fc12577ab00325cf1/\\$file/03_migrate%20your%20mp200_ac450_3bse063520_lr.pdf](http://www05.abb.com/global/scot/scot227.nsf/veritydisplay/3aa1cbc01457c20fc12577ab00325cf1/$file/03_migrate%20your%20mp200_ac450_3bse063520_lr.pdf), accessed: 5 July 2012.
- [21] Deck, B. and Naedele, M. (2004), "IT Security for Utility Automation Systems", 21, rue d'Artois, F-75008 Paris, Available at: <http://www.cigre.org>.
- [22] Singh, B. (2007), *Network Security and Management*, 1st Edition, Prentice Hall of India Publishers.
- [23] Stallng, W. (2006), *Cryptography and Network Security Principle and Practices*, 4th Edition, Pearson Prentice Hall Publishers.
- [24] Albanese, J. and Sonnenreich, W. (2004), *Network Security Illustrated*, 1st Edition, McGraw-Hill Publishers.
- [26] Wang, Y., Gu, D., Xu, J., Du, H. (2009), "Hacking Risk Analysis of Web Trojan in Electric Power System", WISM International Conference on Web Information Systems and Mining, pp. 510-514.
- [27] Khelil, A., Germanus, D., Suri, N. (2012), "Protection of SCADA Communication Channels", *Critical Infrastructure Protection*, Vol. 7130, pp. 177-196.
- [28] Endorf, C., Schultz, E., Mellander, J. (2006), *Intrusion Detection & Prevention*, 1st Edition, McGraw-Hill Publishers.
- [29] Carcano, A., Coletta, A., Guglielmi, M., Masera, M., Fovino, I.N., Trombetta, A. (2011), "A Multidimensional Critical State Analysis for Detecting Intrusions in SCADA Systems", IEEE Transactions on Industrial Informatics, Vol. 7, Issue 2, pp. 179-186.
- [30] Fernandez, E. and Larrondo-Petrie, M. (2010), "Designing Secure SCADA Systems Using Security Patterns", 43rd Hawaii International Conference on System Sciences (HICSS), pp. 1-8.
- [31] Morris, T. and Pavarupu, K. (2010), "A retrofit network transaction data logger and intrusion detection system for transmission and distribution substations", IEEE International Conference on Power and Energy (PECon), pp. 958-963.
- [32] Liu, F., Wang, Q., Gao, X. (2006), "Survey of Artificial Immune System", ISSCAA 1st International Symposium on Systems and Control in Aerospace and Astronautics, Vol. 5, pp. 989.
- [33] Hong, L. (2008), "Artificial Immune System for Anomaly Detection", KAM Workshop IEEE International Symposium on Knowledge Acquisition and Modeling Workshop, pp. 340 - 343.
- [34] Dasgupta, D., Ali, M., Abercrombie, R., Schlicher, B., Sheldon, F. (2011), "Secure VM for monitoring industrial process controllers", CSIIRW '11 Proceedings of the Seventh Annual Workshop on Cyber Security and Information Intelligence Research.
- [35] Ekanayake, J., Liyanage, K., Wu, J., Yokoyama, A., Jenkins, N. (2012), *Smart Grid : Technology and Applications*, John Wiley & Sons Publishers.
- [36] Katz, J. and Lindell, Y. (2008), *Introduction to Modern Cryptography*, Chpman and Hall publishers, ISBN: 9781584885511.
- [37] Cai, N., Wang, J., Yu, X. (2008), "SCADA system security: Complexity, history and new developments", 6th INDIN IEEE International Conference on Industrial Informatics, pp. 569-574.
- [38] Solomakhin, R., Tsang, P., Smith, S. (2010), "High Security with Low Latency in Legacy SCADA Systems", IFIP Advances in Information and Communication Technology, Vol. 342, pp. 63-79.
- [39] Hong, S. and Lee, M. (2010), "Challenges and Direction toward Secure Communication in the SCADA System", (CNSR) Eighth Annual Communication Networks and Services Research Conference, pp. 381-386.
- [40] Bagaria, S., Prabhakar, S.B., Saquib, Z. (2011), "Flexi-DNP3: Flexible distributed network protocol version 3 (DNP3) for SCADA security", (ReTIS) International Conference on Recent Trends in Information Systems, pp. 293-296.
- [41] Lázaro, J., Astarloa, A., Zuloaga, A., Bidarte, U., Jiménez, J. (2011), "I2CSec: A secure serial Chip-to-Chip communication protocol", *Journal of Systems Architecture*, Vol. 57, Issue 2, pp. 206-213.
- [42] AGA (2012), *American Gas Association*, Available at: [http://www.aga.org/our-issues/security/Documents/SCADA%20Protections%20Review%20-%20AGA%20June%202012%20\(Final\).pdf](http://www.aga.org/our-issues/security/Documents/SCADA%20Protections%20Review%20-%20AGA%20June%202012%20(Final).pdf), accessed: 11 July 2012.
- [43] Furht, B. and Escalante, A. (2010), *Handbook of Cloud Computing*, Springer publishers.
- [44] Chang, W. Y., Abu-Amara, H., Sanford, J. F. (2010), *Transforming Enterprise Cloud Services*, Springer publishers.
- [45] Combs, L. (2011), *Cloud Computing For SCADA*, Control Engineering Magazine, Available at: [http://www.controleng.com/index.php?id=483&cHash=081010&tx_tnews\[tt_news\]=63181](http://www.controleng.com/index.php?id=483&cHash=081010&tx_tnews[tt_news]=63181), accessed: 12 July 2012.

e-Monitoring for haul road maintenance in mining applications

PS Heyns

Department of Mechanical and Aeronautical Engineering
University of Pretoria
Pretoria 0002
+2712 4202432
stephan.heyns@up.ac.za

HM Ngwangwa

Department of Mechanical and Industrial Engineering
University of South Africa
PO Box 392, Pretoria 0003
+2711 471 2079
hngwangwa@gmail.com

T Heyns

Council for Scientific and Industrial Research
PO Box 395
Pretoria 0001
+2712 8414788
theyns@csir.co.za

SF van der Westhuizen

Technical Solutions Division of Anglo American
PO Box 61587 Marshalltown
Johannesburg 2107
+2711 638 2046
stephanus.vanderwesthuizen@angloamerican.com

ABSTRACT

Conventional haul road maintenance strategies, including ad hoc blading, scheduled blading and managed maintenance are generally suboptimal for dealing with the complex and rapidly changing conditions encountered in haul road environments. This paper presents an overview of the development and assessment of a two real time haul road condition monitoring techniques, one based on physical models and the other on artificial neural networks. These techniques both rely on an interpretation of the road condition based on the dynamic vehicle response as measured over different sections of road. The thrust of the research is not only to ensure the efficiency of the proposed algorithms, but also to simplify the modelling stage and subsequently reduce the cost for implementing such diagnostic approaches. After comprehensive investigations on the physics based approach at the Exxaro Grootegeluk Mine in South Africa, Anglo American partnered with the University of Pretoria to develop a haul road condition monitoring system based on the artificial neural network approach. This system has since been deployed at three Anglo American mining operations and allows real time monitoring of haul road condition from a central e-monitoring facility.

Keywords

e-monitoring, haul road condition monitoring, mining roads, road maintenance, unpaved roads.

1. INTRODUCTION

Haul road networks in mining applications are typically 10 to 40 km long and comprise various road segments with variable traffic, construction and material characteristics. While maintenance of these roads have in the past relied heavily on human experience, ever increasing truck sizes have resulted in unpredictable road performance, inadequate road maintenance scheduling and excessive operating costs [1]. With haulage costs accounting for 40 to 50% of operating costs incurred by a surface mine [1,2] and routine road maintenance being carried out almost daily, potential cost savings in this area could be very significant.

Four levels of routine haul road maintenance strategies are described in the literature [2]: Ad hoc blading (reactionary maintenance in response to poor haul road functionality), scheduled blading (maintenance according to a fixed schedule irrespective of actual road segment functionality), managed maintenance (road network analysed to determine rate of functional deterioration in individual

segments) and real-time road maintenance (instrumented truck fleet used to assess road functionality based on vehicle response). In the haul road context where road deterioration can occur very rapidly, the use of on-board data-collection systems in conjunction with standard global positioning systems (GPS) and radio communication systems has been proposed for haul road maintenance in opencast mining applications.

Road condition monitoring can be accomplished directly by visual inspection, profilometer, laser technology, ground penetrating radar, etc. Here, however, we consider an indirect approach based on measured vehicle response to identify road defect signatures, as was conceptually suggested by Thompson et al. [3]. Such an indirect approach offers significant advantages in respect of simplicity of instrumentation for real time e-monitoring, but suffers from the problem that the measured response signal is not directly related to the road condition alone. In addition to the effects of road condition, measured vehicle response is also highly sensitive to vehicle operating speed, vehicle load, and the unique vehicle characteristics.

In this paper we consider the specific problem of reconstructing road profiles based on measured truck response, and then using these road profiles for monitoring purposes. To implement such an indirect approach to establish the road profile condition, one could rely on physical modelling of the truck. The inverted truck model may be used to estimate the road profile from the measured vehicle response [4,5]. The spectrum of the estimated road profile may subsequently be interpreted against the ISO road classification standards. While it is indeed possible to accurately model the whole vehicle, this approach is both complex and expensive. Hugo et al. [4,5] demonstrate through physical experiments at the Grootegeluk Mine in South Africa, that the measurement and the experimental component characterisation process may be simplified, as well as the problem of inverting the model largely avoided, if the haul truck model is reduced to a very simple single-degree-of-freedom model of a front wheel and suspension. This single-degree-of-freedom model renders satisfactory results as is demonstrated, but characterisation of the highly nonlinear tyre and large scale strut is still a difficult and expensive process.

It is therefore subsequently investigated whether an artificial neural network (ANN) may be used to represent the inverse dynamics of a haul truck [5]. This avoids the need to individually characterise vehicle components, by allowing the ANN to learn the vehicle characteristics from a set of training data. The training data comprise known road profiles and corresponding measured vehicle responses.

As previously discussed the vehicle response is not only affected by the road condition, but also by the operating conditions of the vehicle. This is a challenging and very relevant problem which has recently been further explored in the context of robust statistical approaches [7,8]. For the purposes of the work reported here, it is assumed that the problem of fluctuating operating conditions can be dealt with adequately by appending the input vector in the training data set with measurements of the instantaneous operating conditions. The ANN approach renders it possible to model non-linear characteristics of the vehicle, such as those exhibited by tyres and suspension systems. The approach is first demonstrated using simple numerical simulation models as well as real experiments on a sport utility vehicle. The basic ANN approach was subsequently successfully applied to a fleet of haul trucks at three other mines. This implementation for haul road maintenance is also briefly presented here.

2. ROAD PROFILE RECONSTRUCTION BASED ON A PHYSICAL MODEL

For the physical model based work a Komatsu 730E Haulpack was considered by a University of Pretoria team at the Exxaro Grootegeluk Coal Mine. In this work a 7DOF rigid body vehicle model is used to approximate the truck dynamics. The model allows for vertical, pitch and roll motion of the sprung mass, vertical and roll motion of the rear unsprung mass and vertical motion of the two independent front unsprung masses. The characterised model in conjunction with the measured vehicle response data indicate that the essential dynamics of the vehicle can be captured well in a 0 to 10 Hz range. The suspension strut forces can be estimated quite accurately using the 7DOF model.

A circular rigid treadband tyre model is assumed [4,5]. This model represents the tyre as a simple spring that is attached to a rigid ring that represents the treadband of the tyre. This is equivalent to filtering the road profile by effectively lowering the treadband from a vertically elevated position until one point on the treadband makes contact with the road and the corresponding elevation of the centre of the treadband may be determined. Using the 7DOF model in conjunction with measurement results and a thorough physical understanding of the vehicle response, it is subsequently shown that a simple single-degree-of-freedom model of the independent front sprung mass provides an adequate description of the vehicle response dynamics.

Figure 1 depicts such a rigid free-body diagram of the unsprung front mass. Assuming the acceleration of the unsprung mass \ddot{z}_u and the force acting between the sprung and unsprung masses F_{susp} to be measured as functions of time, the force exerted on m_u by the tyre F_{tyre} can be determined from the very simple algebraic equation:

$$F_{tyre} = m_u \ddot{z}_u + F_{susp} \quad (1)$$

Because of the non-linear spring and damping characteristics of the hydro-pneumatic suspension struts, measuring anywhere on the sprung mass would generally require the inversion of a non-linear system in order to accurately calculate the road input from the measured response. In the simple model proposed, the

suspension strut force is measured and hence known. Its non-linear dependence on relative suspension strut displacement and velocity therefore does not feature in the solution, and neither is the inversion of a full non-linear system required. Furthermore, characterization of the complete vehicle is not necessary.

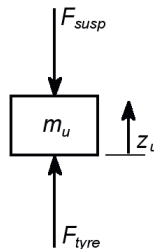


Figure 1 Simple SDOF model on front unsprung mass

However excitation due to tyre/wheel non-uniformities cannot be distinguished from the tyre force F_{tyre} . This is however not considered to be very important for the large tyres and rough terrain considered in the typical mining haul road application. Once the tyre force F_{tyre} is available, the unsprung mass displacement and an inverse tyre model are required to calculate the corresponding road input.

As an alternative to inverting the vehicle model, the road profile can also be reconstructed through optimization. Dynamic equilibrium is enforced on the independent front unsprung mass. Several types of road input are provided to such a simulation model, in which F_{susp} is treated as a known variable. The road profile can be determined by iteratively changing the road input until the simulated unsprung mass acceleration matches the measured unsprung mass acceleration. It should be possible to use tyre models with different levels of complexity in such a model, with the advantage that inversion would not be required.

2.1 Component characterisation

Solution of the vehicle response requires characterization of the tyre and strut dynamic properties. The tyre characteristics were determined in situ on an open-air weighbridge. The truck was loaded with a range of coal loads while measuring the weight on each wheel. At the same time the tyre deflection was also recorded. From this an equivalent linear stiffness coefficient of 2713 kN/m was obtained. Tyre damping was disregarded.

The front suspension struts simultaneously provide the stiffness and damping at the front of the vehicle. Damping is caused by oil flowing through ports in the side of the suspension rod. During extension some of the damping ports are closed by valves. Consequently, damping becomes higher during extension than in compression. The strut also contains nitrogen gas that provides the spring force in the strut. The nitrogen lies on top of the oil, with the oil level high enough to cover the damping ports. This implies

that the strut must be operated in an essentially vertical orientation. The strut was mounted vertically against a reaction wall with a 630 kN actuator being used to dynamically load the strut at frequencies ranging from 0,005 Hz to 0,3 Hz. This provided spring force as a function of displacement for cases that respectively approached isothermal and adiabatic conditions.

To obtain the damping characteristics the strut was forced with saw tooth displacement signals over a range of frequencies and the total suspension strut force was recorded (with a load cell), the strut displacement was measured using the actuator LVDT (and differentiated to obtain velocity) and the nitrogen gas pressure was measured used to obtain the suspension strut spring force. Because of the very large forces involved, the damping forces could only be measured up to actuator speeds of about 0,06 m/s while data was required to significantly higher speeds. For this reason a thermo-fluid model of the suspension characteristics was developed and very good correlation between theory and experiment was found up to 0,06 m/s. This proved that the basic underlying theory was well understood and the model was subsequently used to extrapolate the experimental data.

2.2 Road defects

Treating the tyre like a simple linear spring effectively allows inversion through

$$z_r = z_u + \frac{F_{tyre}}{k_t} \tag{2}$$

where z_r is the road elevation and z_u is the unsprung mass displacement. F_{tyre} may be calculated from Equation (1), in which the measured unsprung mass acceleration is used directly. F_{susp} is calculated from the characteristics of the suspension strut. The only unknown remaining in Equation (2) is the unsprung mass displacement. As it is difficult to measure the absolute displacement, the obvious alternative is to integrate the unsprung mass acceleration twice.

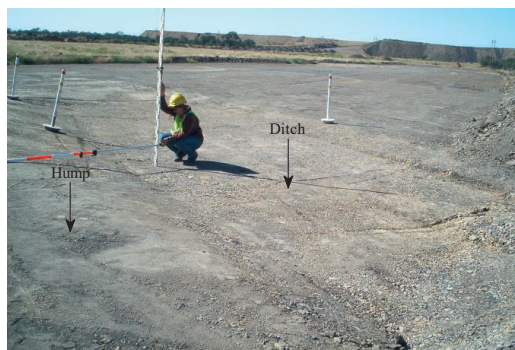


Figure 2 Large ditch followed by hump

The measured unsprung mass acceleration was therefore integrated twice in the frequency domain with no lower cut-off frequency, and used in Equation (2). Using the known speed of

the vehicle, the resulting profile was transformed to the spatial domain. Because of the problems associated with low-frequency integration, severe drift was however still present. This drift was removed by fitting third-order polynomials to selected portions of the profile, in such a way as to approximate the low frequency drift but not the defect itself. The curve described by this polynomial was then subtracted from the profile obtained from using Equation (2). By way of example, figure 2 depicts a defect that was constructed in this way. Figure 3 visually compares typical nominal and vibration based estimated road profiles.

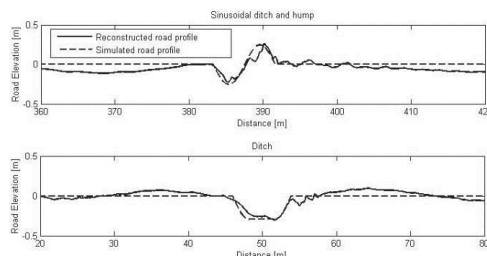


Figure 3 Reconstructed defects compared to nominal geometry

3. ROAD PROFILE RECONSTRUCTION BASED ON ARTIFICIAL NEURAL NETWORKS

3.1 Exploratory numerical investigation

While physical model based road profile reconstruction was proved successful at the Grootegeluk Mine, the expense and practical difficulties of characterising the tyres and suspension struts limits implementation of this approach in the field. To circumvent these problems an alternative approach was numerically explored by Ngwangwa, Heyns, Labuschagne and Kululanga [6]. In this approach an artificial neural network is used to reconstruct road surface profiles from measured vehicle accelerations.

To investigate the feasibility of the methodology, a simple 8DOF numerical linear pitch plane model with four degrees of freedom representing vehicle motions and additional 4DOFs for seat and driver responses was first employed to generate response data. Eight roughness grades having well-known road profiles on the ISO PSD classification [6] were applied to the model to calculate corresponding sprung mass and axle accelerations. The generated training data were subsequently used to train the neural network.

The input to the ANN comprises a vector which contains the lumped body accelerations, while the corresponding road profiles serve as regression targets. An auto-regressive with exogenous inputs (NARX) ANN topology was implemented which comprise 20 neurons in the hidden layers with tansigmoid activation functions, and one linear output corresponding to the instantaneous road elevation.

Once the ANN was trained a second set of evaluation data was generated, which were used to validate the ability of the ANN to

generalize well on novel data. The simulated road profiles are compared with the actual road profiles. These evaluations were performed for different road roughness conditions, different vehicle speeds, growing road defects, different noise levels, and different vehicle payload conditions.

These exploratory numerical investigations provided sufficient justification for full scale field exploration. This was first done on a sport utility vehicle (not reported here) and then on a haul truck.

3.2 Field measurements

Based on the encouraging results obtained with the preliminary studies [6], a collaborative project between the University of Pretoria and Anglo American was conducted on a haul truck at Mogalakwena Mine in South Africa. This truck is depicted in figure 5.



Figure 5 Haul truck at Mogalakwena Mine

The truck was instrumented (see figure 6) with accelerometers, a GPS and an optical sensor which is sensitive to reflective strips which were placed on the ground, for relative vehicle-ground positioning. The instrumentation list is summarised in Table 1.

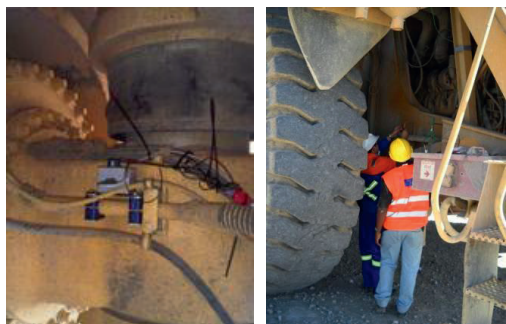


Figure 6 Typical instrumentation set-up

Table 1 Mogalakwena Mine instrumentation list

Measurement quantity	Sensor	Measurement point
Time	From eDAQ	
Left Front Wheel Acceleration in vertical Z direction <i>[LFWheelAcc]</i>	Tri-axial 4g accelerometer	Spindle of the left front wheel
Right Front Wheel Acceleration vertical Z direction <i>[RFWheelAcc]</i>	Tri-axial 4g accelerometer	Spindle of the right front wheel
Rear axle acceleration <i>[RearAcc]</i>	Tri-axial 4g accelerometer	Off axle centre on any side that is more convenient
Vehicle speed <i>[VehSpeed]</i>	GPS	
Vehicle position <i>[VehPosition]</i>	Position probe	Anywhere on vehicle provided the distances from the other measurement points are specified

More test details are provided in table 2.

Table 2 Test details

Road measurements	Test road length: 50 – 100 m. Length between adjacent samples: 4 – 5 m
Sampling frequencies	Response data at 400 Hz
Anti-aliasing filter: Cut-off frequencies and acceptable phase distortion	Response data: 200 Hz cut-off with a 50Hz roll-off

3.3 Artificial neural network

Various artificial neural network architectures were investigated for this application. A typical network is depicted in figure 7.

The artificial neural network operates by regressing outputs and inputs, from previous fifteen time steps, on the current output $\tilde{z}_r(t)$, thus

$$\tilde{z}_r(t) = f(\tilde{z}_r(t-1), \tilde{z}_r(t-2), \dots, \tilde{z}_r(t-15), p(t-1), p(t-2), \dots, p(t-15))$$

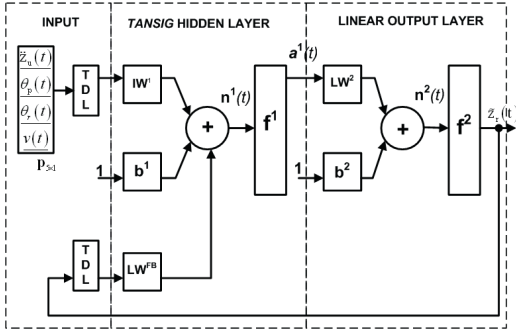


Figure 7 ANN architecture

The ANN in its architecture is given by:

$$\tilde{z}_r(t) = f^2[LW^2 * f^1(IW^1 * \{p\}TDL + b^1 + LW^{FB} * \{\tilde{z}_r\}TDL) + b^2]$$

where f^1 is a tansigmoid function, f^2 is a purelin function and superscripts TDL and FB denoted 15 time delay lines and feedback respectively. Thus we have the function rewritten

$$\tilde{z}_r(t) = \text{purelin}(LW^2 * \text{tansig}(IW^1 * p(t_{-1}, t_{-2}, \dots, t_{-15}) + b^1 + LW^{FB} * \tilde{z}_r(t_{-1}, t_{-2}, \dots, t_{-15})) + b^2)$$

where input weights IW , layer weight LW , and the biases $\{b^1, b^2\}$, are final coefficients from a trained network.

During training, iterated computations of $\tilde{z}_r(t)$ are carried out until the sum of squared errors E_D and network weights E_w are minimized until a predefined criterion is satisfied. The optimized weights and biases may subsequently be used to infer the road profiles which correspond to newly acquired vehicle responses.

3.4 Practical implementation

Anglo American developed a haul road condition monitoring system. This system implements the artificial neural network approach to profile reconstruction as is described in paragraphs 3.2 and 3.3.

Instrumented haul trucks measure the condition of the haul roads under operational conditions. A central server receives data from three to five trucks at each operation. The haul road condition data is presented via an interactive web interface, which can be

accessed from a centralised monitoring facility in Johannesburg or directly at each deployment site.

The objective of the project is to allow both centralised and on-site real time monitoring of the haul road condition at all Anglo American opencast mining operations to improve haulage performance and to identify value opportunities.

The system has, to date, been implemented at three Anglo American operations and useful results are being obtained.

4. CONCLUSIONS

It has been shown that response based e-monitoring can be applied successfully for haul road maintenance in mining applications. Two approaches have been explored in actual haul road applications. In the first approach physics based road profile reconstruction is conducted. While good results are obtained, the approach is expensive because of the very large scale characterisation experiments that need to be conducted. This remains the case despite successful endeavours to reduce the original 7DOF physical model to a very simple single degree of freedom model of the front unsprung mass.

For this reason the use of artificial neural networks for road profile reconstruction is first numerically investigated on an 8DOF vehicle model and then practically implemented on an actual haul road truck after experiments at Mogalakwena. This proves the usefulness of the system which was subsequently also implemented at two other Anglo American operations.

Dealing with the effects of varying operating conditions in this approach remains challenging. The use of sophisticated statistical techniques to deal with these issues have therefore been investigated in trial runs on utility vehicles at Anglo American mines, but remains to be implemented on haul trucks.

5. REFERENCES

- [1] Hugo, D., Heyns, P.S. Thompson, R.J. and Visser, A.T. (2007). Condition triggered maintenance for mine haul roads with reconstructed-vehicle response to haul road defects. Transportation Research Record: Journal of the Transportation Research Board, No.1989, vol.2, Transportation Research Board of the National Academies. Washington, D.C., 254-260.
- [2] Nebot, E.M. (2007). Surface Mining: Main research issues for autonomous operations. Robotics Research Results of the 12th International Symposium ISRR, Edited by Thrun, S., Brooks, R. and Durrant-Whyte, H.
- [3] Thompson, R.J., Visser, A.T., Millar, R.E. and Lowe, N.T. (2003). The development of a real-time mine road maintenance management system using haul truck and road vibration signature analysis. The Journal of the South African Institute of Mining and Metallurgy, June 2003, pp.265-271.
- [4] Hugo, D., Heyns, P.S. Thompson, R.J. and Visser, A.T. (2005). Haul road condition monitoring using vehicle response measurements. Twelfth International Congress on Sound and Vibration, 11-14 July 2005, Lisbon.

- [5] Hugo, D., Heyns, P.S. Thompson, R.J. and Visser, A.T. (2008). Haul road defect identification and condition assessment using measured truck response. *Journal of Terramechanics*, vol.45, 79-88.
- [6] Ngwangwa, H.M., Heyns, P.S. Labuschagne, F.J.J and Kululanga, G.K. (2010) Reconstruction of road defects and road roughness classification using vehicle responses with artificial neural networks simulation. *Journal of Terramechanics*, vol.47, pp.97-111.
- [7] Heyns, T, Heyns, P.S. and De Villiers, J.P. (2012) A method for real-time condition monitoring of haul roads based on Bayesian parameter estimation. *Journal of Terramechanics*, 49, 2012, pp.103-113.
- [8] Heyns, T, De Villiers, J.P. and Heyns, P.S. (2012). Consistent haul road condition monitoring by means of vehicle response normalisation with Gaussian Processes. *Engineering Applications of Artificial Intelligence*. Doi:10.1016/j.engappai.2012.01.009.

CMMS Benchmarking Development in Mining Industries

Yonas Lemma
Operation and
Maintenance Eng.,
Luleå University of
Technology, Sweden
yonlem-0
@student.ltu.se

Stephen M.
Famurewa
Operation and
Maintenance Eng.,
Luleå University of
Technology, Sweden
stefam@ltu.se

Diego Galar
Operation and
Maintenance Eng.,
Luleå University of
Technology, Sweden
diego.galar@ltu.se

Håkan
Schunnesson
Geotechnology
Luleå University of
Technology, Sweden
hakan.schunnesson
@ltu.se

Jonas Fjellner
Boliden AB
Boliden, Sweden
jonas.fjellner
@boliden.com

ABSTRACT

Effective resource management and reliable equipment are essential to optimize plant performance. Both depend on accurate and timely management of massive amount of data and on the effective use of maintenance resources. Computerized management system (CMMS) also called computerized asset management system (CAMS), are designed to fulfil these needs. This system can support cost effective means of managing and utilising a massive amount of data that are generated by maintenance, inventory control, operation, purchasing and other relevant activities. In addition, this system can provide the means to manage effectively both human and capital resources in plant or facility operation. However to achieve the intended purpose of the system it is essential to assess its utilisation to identify functional gaps. This will involve total evaluation of different aspects of CMMS function called criteria. This paper presents the challenges of successful CMMS implementation and also describes CMMS benchmarking process. A case study of CMMS benchmarking in two underground mine locations of a mining company are presented. The result from the questionnaire used in the investigation is compared with 33 paper mill factories in different European countries using the same computer maintenance management system software. The contribution of this work can be summarized as follow: Identifying relevant measures that can be used to quantify functional gap in implemented CMMS system and using CMMS benchmarking procedure to suggest improvement towards maintenance excellence.

Keywords

Benchmarking, Computer maintenance management system, Maintenance performance measurement

1. INTRODUCTION

Maintenance can be defined as the orderly control of activities required to keep a facility in an as-built condition, with the ability to sustain its original productive capacity (1). The management of maintenance functions is involved with the following tasks; determine the objectives, strategy, and does other responsibilities such as planning, control, supervision and improvement of the function (2).

A good maintenance management system makes equipment and facilities available. During the past 20 years, the term "CMMS" (computerized maintenance management system) has become

synonymous with productivity improvement and control of maintenance management processes. In the last three years, literally billions of dollars has been spent in many different countries, on implementing CMMS and enterprise level systems. Some corporations have attempted to circumvent a lot of this cost by developing implementation templates for use across their global operations (3).

Maintenance prognosis and optimal decision making requires effective integration of disparate data sources including, production, failure, preventive maintenance, condition monitoring, inventory and other relevant asset information sources. This is possible through the implementation of a CMMS. A CMMS with a seamless integration of CM system can perform a wide variety of functions to the advantage of asset owners. It brings dramatic results in asset reliability, cost effectiveness and maintenance process efficiency (4). This tool is indispensable for maintenance improvements.

CMMS implementations follow basically the same principle every time. Yet most times there is an attempt to re-invent much of the approach. However, in general maintenance is apart from differing business rules the majority of the information required to set up the system will remain similar. However the need for CMMS benchmarking cannot be undermined as it is an important tool to help gain maximum value from an existing one or from the implementation of a new system. The CMMS benchmarking system is introduced as a means to evaluate the effective use of any current CMMS, to define functional gaps and to suggest how to enhance current use. Results from a well-designed CMMS benchmarking system will also help to develop and justify a replacement strategy if that is needed. It can also be used as a method to measure the progress of a CMMS system implementation that is now being installed and its future success (5).

The objective of this paper is to study and evaluate the implementation and contribution of CMMS in mining industry. The following research questions have been formulated to achieve the set purpose and objectives of this study;

- What are the relevant measures that can be used to assess the performance of an implemented CMMS system?
- How can CMMS be effectively used as a tool to support best practices and maintenance excellence?

The structure of the article is as follows: section II gives the description of the research methodology and section III describes some aspects of CMMS benchmarking process. A case study is presented in section IV thereafter the result and discussions are presented in sections V and VI respectively. A concluding remark is given in the final part of the paper.

2. RESEARCH METHODOLOGY

The approach of this research is both literature study and survey. Literature related to maintenance management system was studied and the essentials of both CMMS and benchmarking are synthesised from the literature. The research study was carried out using the survey instrument, questionnaire. The questionnaire followed the standard approach commonly used for CMMS benchmarking in order to assess the CMMS implementation and valuable contribution of this tool to maintenance objectives. The quality and quantity of the historical data from the studied CMMS was also evaluated.

The questionnaire was sent to persons having responsibility related to maintenance in the case study area. The target audience include maintenance director, manager, supervisor and shop floor technician. A group response was given to the questionnaire. The detail of the response and relevant analysis are given in later section. The investigation through questionnaire helps to gather information about the CMMS implementation in the case study area and to further explore associated problems and success factors. This would help to find answers to the research questions and to understand the topic at depth.

The design of the questionnaire is in such a way to reflect the important aspects of CMMS function. After identifying the standard functional requirements from literature survey (6), (5), (7), questions were formulated to measure the achievements of these requirements by the implemented CMMS in the studied underground mines. Other relevant detail of the questionnaire is giving in later section.

3. THE CMMS BENCHMARKING SYSTEM

Why use benchmarking? A basic truth is measurement is a prerequisite to analysis, knowledge, control and effective management. This informs the use of benchmarking to measure what is to be managed. Benchmarking is a very versatile tool that can be applied in a variety of ways to meet a range of requirements for improvement (5). Benchmarking is an effective management tool to identify gaps in the way an existing activity, function, or process is performed. These are change ideas that when implemented continuously helps to achieve continuous improvements. In employing this method, a company compares its performance with its strong and more successful competitors in the industry. It helps a company not only assess its current performance relative to other companies, but also learn from others and generate new ideas, methods and practices to improve its functioning. Thus, productivity and cost reduction can be enhanced and new performance targets, which are practical and achievable, can be set to give itself a competitive edge (8).

Benchmarking is required to improve the management of facility and the return on such assets in terms of its output and

performance. In most cases, the higher the quality of maintenance being performed, the greater the return on assets will be. Therefore, to optimize the management of a physical asset or return on the assets, it would be valuable to determine where the relevant performance measure compares to the rest of the industry through the use of benchmarking procedure (9). Many people see benchmarking only as a method for comparing key figures very often financial or as tool for ranking a company against competitors, conversely it extends beyond these. The purpose of a benchmarking study is not only comparing for the sake of evaluation but learning for achieving excellence and best practices. One does not compare only financial figures, other non-financial performance measures are important element in the process, i.e. It extends to how tasks are performed or how tools are utilised; gives insight into how those who are better develop and implement their processes. The learning effects are not limited to information available from competitors. It is rather encouraged to take an external view, seeking the best companies regardless of industry. Benchmarking is not some left hand task that one hires a consultant to do. Benchmarking should be done according to a structured process, where one self harvests the learning effects (10).

If any problem areas exist, benchmarking can detect them, and then those problem areas can be removed or minimized.

The CMMS benchmarking system is a means of evaluating the effective use of the current CMMS implementation to define functional gaps and to define how to enhance current use. CMMS benchmarking is anticipated to measure its level of implementation, additional value to business goal, practices, and develops its own distinctive benchmarking criteria with high standards for maintenance excellence. An Effective CMMS benchmarking should start at the initial stage of mapping the maintenance organisation, current maintenance practices and procedures (5). These ultimately inform the utilisation and functional requirements of CMMS.

4. CASE STUDY

The case study is carried out in two underground mine operations in Sweden. Both mines use Maximo asset management software as maintenance management system. This research work focusses on the evaluation of the effectiveness of current CMMS utilization by using generic benchmarking tools. The study also extends to comparison of the result with different paper mill companies, using the same type of CMMS for maintenance management. The first task was to map the maintenance process and organisational structure of the case study. Maintenance organization describes the interaction of inputs and outputs in a maintenance process. This is characterized by assignment of tasks, communication channels, reporting procedures and workflow that link together the activities of different individuals and groups within the process (11). The organisational structure of maintenance affects the implementation of CMMS to a large extent. The mapping of the maintenance organisation structure in the case study is shown in Figure 1. This gives a preliminary understanding of the set up and functional linkages of maintenance activities with the reporting procedure as well.

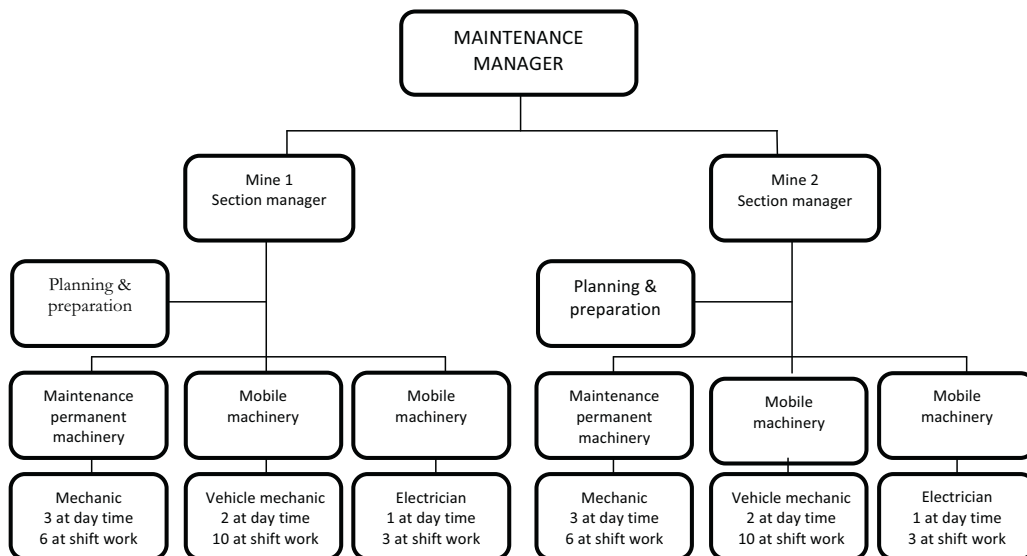


Figure 1. Maintenance organization of the two underground mines

5. RESULTS

This section presents the results, analysis and findings from the benchmarking study of the two underground mining sites. The implementation and utilization of CMMS is understudied and compared with different paper mill factories in Europe. Following the benchmarking model of (10), we have proceeded in the CMMS benchmarking of the organization in the case study.

As stated in one of the research questions, this study intends to identify relevant criteria that can be used to assess the performance of an implemented CMMS system in mining industries. The performance aspects of CMMS, which are considered significant in this study, are listed in Table 1. These evaluation criteria are suggested by (5), and are used to develop a questionnaire. Useful information are collected from the questionnaire using benchmarking rating and presented in a logical form scale in table 1.

To extract knowledge from the information presented in table 1, visual presentations are done in graphs and comparisons are made by generic benchmarking procedure.

A column chart shown in figure 2 is developed to compare the average status of 33 paper mill factories and underground mine 1 & 2 concerning results of same CMMS evaluating categories. This is an external benchmarking procedure and it identifies some weak aspects of CMMS implementation. The major weak areas in the CMMS implementation of the case study are; CMMS education and training, budget and control aspects, preventive and predictive maintenance aspects and maintenance performance aspects. However, the mines have laudable use of CMMS in material management and planning & scheduling.

Table 1. Description of the criteria in the questionnaire and the number of questions in each with respective points

CMMS BENCHMARK EVALUATION CATEGORIES	EVALUATION ITEMS	OUT OF POINTS	SWEDEN (Mine 1 & 2)
1.CMMS Data Integrity	6	24	15
2.CMMS Education and Training	4	16	0
3.Work Control	5	20	10
4.Budget and Cost	5	20	6
5.Planning and Scheduling	7	28	19
6.MRO material Management	7	28	12
7.Preventive and Predictive maintenance	6	24	3
8.Maintenance Performance Measurement	4	16	4
9.Other Uses of CMMS	6	24	9
Total CMMS evaluation Items	50	200	78

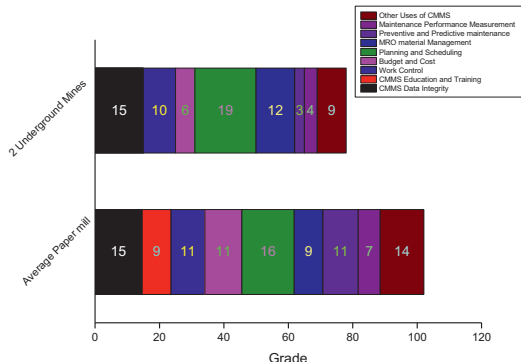


Figure 2. CMMS Benchmarking Rating: comparison between mining and paper mill companies.

Further analysis is done to compare the performance of some CMMS implementation in different countries. This is shown in Figure 3, based on the information available in this study, the two underground mines in Sweden falls in an average position ahead of two other countries and behind the remaining two. Not so much could be inferred on cultural influence on the effectiveness of CMMS implementation in mines.

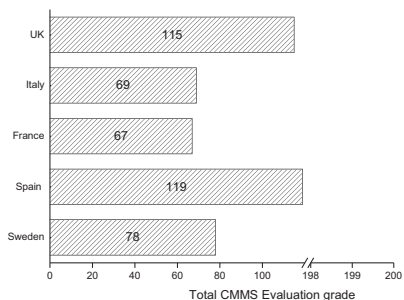


Figure 3. CMMS Benchmarking Rating or total grade: comparison between different countries

Figure 4 compares the total grade of the benchmarking procedure for the two underground mines with the average, worst and best of some paper mills with similar CMMS. The two underground mines fall below the average grade of the 33 paper mills; this is reflecting a potential room for improvement of the CMMS implementation in the mines.

To further investigate the implementation of the CMMS in use in the underground mines, data are collected from the Maximo asset management software to check the status of the key performance indicators (KPI) of the maintenance function. It is also intended to check the lesson learnt from the system of indicators as facilitated by the CMMS.

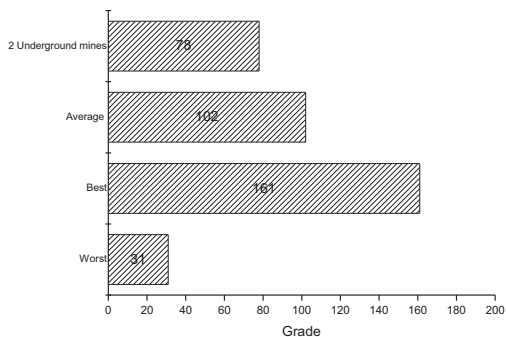


Figure 4. CMMS Benchmarking: rating and comparison between a mining company and selected paper mill companies

KPI development and monitoring is one of the important outcomes that you get from a CMMS (7). It is essential that you assess this to improve your CMMS implementation, as well as redefine your maintenance performance framework to achieve the required organisational objectives. CMMS can facilitate continuous improvement of maintenance function using the key performance indicators. The different aspects of maintenance performance that can be assessed and monitor by CMMS include financial, equipment performance, learning and growth, safety, quality, other maintenance process aspects.

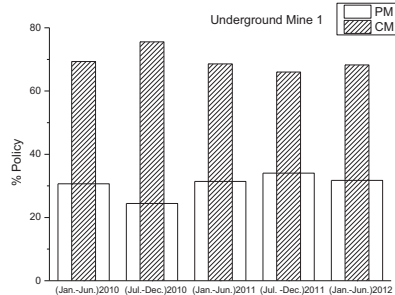


Figure 5. Preventive maintenance vs corrective maintenance underground mine1 from CMMS data

Some aspects of maintenance performance related to CMMS implementation in the mines are studied. Figure 5 and 6 show the level of preventive and corrective maintenance over three years in the underground mine1 and 2 respectively. The figures indicate a high level of corrective maintenance in the two mines. Further discussion on these is given later in the paper.

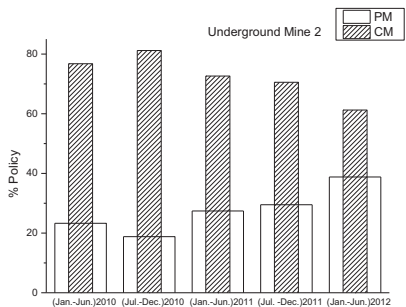


Figure 6. Preventive maintenance vs Corrective maintenance in underground mine 2 from CMMS data

Furthermore schedule compliance of work orders is studied. Schedule compliance is the ratio of the total number of work orders completed as scheduled divided by the total number of work orders that was scheduled. It is a measure of adherence to the maintenance schedule. This KPI was computed over a period of six months and the result is presented in Figure 7 to compare the two underground mines. The figure shows a good compliance to schedule of maintenance services but the plan estimation accuracy is another KPI that is needed to establish the status of schedule compliance.

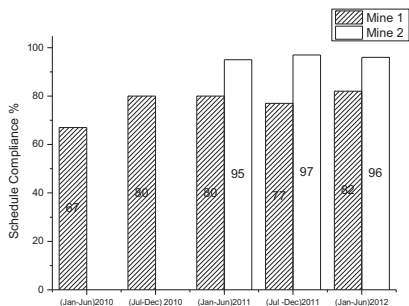


Figure 7. Schedule Compliance underground mine 1 & underground mine 2

Figure 8 shows the data quality with respect to scheduled and actual maintenance time. This is an indication on how essential maintenance data are filled in the CMMS. Further discussion on this is done later.

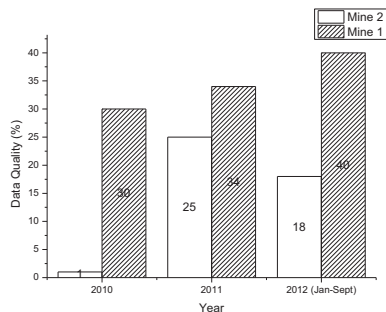


Figure 8. Mine 1 and mine 2 reliable data for Schedule and actual hours

Another KPI, which is an outcome of CMMS, is plan estimation accuracy. It is a ratio of actual number of labor hours reported on a work order to the estimated number of labor hours that was planned for that work order.

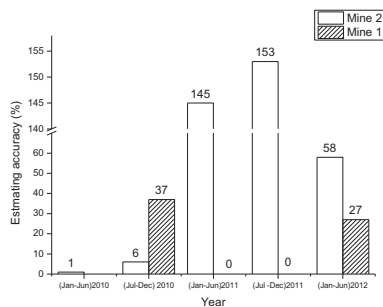


Figure 9. Actual hours to plan estimate, underground mine 1 & underground mine 2

6. DISCUSSION

The result of the benchmarking of the two underground mines under investigation has shown the pitfalls and potential areas of improvement of the implemented CMMS. To harness and consolidate the potential contribution of the implemented CMMS in the two underground mines to maintenance excellence, the following are considered essential;

CMMS education and training: Referring to figure 1, the result has shown that the mining company has given no attention to training their employee on CMMS implementation. Learning and growth is an indispensable prerequisite to success, which is missing in the company. It is highly recommended that competency based training be conducted so that each person trained can demonstrate competency in each function they must perform on the system (5). It is recommended that the mines should organise initial and on-going training for both maintenance personnel and operators. On a higher level of skill improvement, it is important that each site have one person trained and dedicated as the systems administrator.

Preventive and predictive maintenance

A vital aspect of CMMS is to support maintenance at the both strategic and operational levels (12). An excellent implementation of CMMS should support maintenance activities such as planning and scheduling in a proactive way. The result of the benchmarking shows that the implementation CMMS in the underground mines does not adequately utilise the tool to promote both preventive and predictive maintenance. Update and review of preventive maintenance activities could be guided by the CMMS and also schedule compliance of planning and scheduling can be improved. To harness the potential of CMMS it is necessary to improve task description of preventive and predictive maintenance. This will help new craft person at operational level and also engineers at the tactical levels to make maintenance more efficient and effective.

Data Quality

Data quality issues are critical to the success of asset management. The result of the benchmarking has highlighted some data quality problems, which existed in the current CMMS implementation in the two underground mines. Data quality issues need to be widely understood and managed in order to ensure effective asset management. The quality of data limits consequent analysis, knowledge and decision support. There are certain factors that influence data quality when physical asset management. The factors include: systems integration, training, management support, employee relations and organisational culture, such factors should be sufficiently addressed. Understanding these key factors should lead to high-level data quality management practices, which is a key to the successful implementation of computerised asset management system. Among the essentials maintenance data needed for maintenance analysis, control and improvement are scheduled start and finish time and also actual start and finish time. The contextual measure of quality of data in this case study is the reliability (measure of usefulness) of the mentioned maintenance data.

Adequate training is essential for all personnel involved in asset maintenance especially at the shop floor level where most of the data collection is done. This is important for ensuring and improving data quality. People's ability to use the system is equally important to ensure a relatively high level of data quality in asset management.

Budget and Cost Control

Without an effective work order system, it is difficult to manage and control the total maintenance cost. CMMS helps to record both direct and indirect costs incurred on maintenance. History of all cost related to maintenance such as labour, equipment, vendor support, spares, etc. are stored in a useful format in the CMMS. Cost improvement through maintenance optimization and LCC analysis is supported by CMMS. Many organizations often fail to have access to accurate equipment repair costs to support effective replacement decisions and continue to operate and maintain equipment beyond its economically useful life (5). The result shows that the two underground mines have low utilisation of the CMMS for budget and cost control compared to the paper mills. This is a necessary improvement potential towards a cost effective and world-class maintenance service.

Furthermore, an effective implementation of CMMS will support continuous improvement of maintenance function with the system of indicators. The preventive maintenance (PM) level for the two mines is about 30% over the 2year period investigated in this study, see Figures 5 & 6. This is very low and it reflects low utilisation of the planning and scheduling capability of CMMS.

Also combining information in Figure 7 and 9, it is seen that Mine 2 has better schedule compliance than Mine 1 but the estimation accuracy of maintenance time for the two mines are not satisfactory. Effective implementation of maintenance management system can support the extraction of useful information or lesson learnt from historical data. For example, the distribution of repair time for a component in the past can help predict the length of time it will take to repair similar component in the future.

Finally, internal benchmarking of the data quality aspect of CMMS implementation shown in Figure 8 indicates that mine 1 has better implementation than Mine 2. However both Mines have low record of data quality and this presents opportunity for improvement as discussed earlier on.

7. CONCLUSIONS

Maintenance management requires a tool (CMMS) to effectively gather, monitor and assess data on condition of asset and also relevant maintenance information to facilitate decision-making. The evaluation and benchmarking procedure of such tool has been described in this work. The nine evaluation categories mentioned in this study are relevant performance aspects of an implemented CMMS system, which should be assessed during benchmarking. The analysis of the questionnaire shows that the case study has weakly utilized their CMMS to some extent, according to CMMS benchmarking rating (5). CMMS implementation of the case study falls into Class D, which implies very low utilization, but there is large room for improvement in some aspects of the CMMS implementation. The aspects to be improved include data quality, maintenance performance measurement, CMMS education and training, preventive and predictive maintenance, budget and cost control. Furthermore, assessment of maintenance decisions and actions is a key task in maintenance process, it is essential that CMMS facilitates this so as know the value contribution and deficiency of maintenance. The status of some maintenance KPI in the case study indicates that maintenance performance measurement using CMMS is inadequately deployed. CMMS can support continuous improvement using the system of indicators if its capability is sufficiently harnessed. Finally, the identified improvement areas in the CMMS implementation of the two underground mines are essential requirements towards achieving maintenance excellence.

8. ACKNOWLEDGMENT

The authors wish to appreciate the financial support of Boliden AB and the technical support of Mr. Thomas Sundqvist and Mr. Arne Vesterberg for provision of the data.

9. REFERENCES

- [1] Bagadia, K. (2006). Computerized Maintenance Management Systems Made Easy. McGraw-Hill.

- [2] European Committee for Standardization- CEN, (2007), EN-13306; Maintenance terminology, European Standard, Brussels, 2007.
- [3] Mather, D. (2009). Developing CMMS Implementation Templates. Plant Maintenance Resource Center.
- [4] Galar, D., Palo, M., Van Horenbeek, A., Printelon, L. M. (2012). "Integration of disparate datasources to perform maintenance prognosis and optimal decision making". *Insight - Non-Destructive Testing and Condition Monitoring*, Volume 54, Number 8, pp. 440-445(6).
- [5] Peters, R. (2006). Maintenance benchmarking and best practices. Raleigh, NC-Oak Island, NC: Maintenance Excellence Institute.
- [6] Cato, W. W., and Mobley K. R. (2001). Computer Managed Maintenance System. Elsevier Science & Technology Books.
- [7] Mather, D. (2002). CMMS A Timesaving Implementation Process. Boca Raton London New York Washington, D.C. CRC PRESS.
- [8] Wheelen, L., Thomas, DHJ, Krish, R. (2006). Strategic Management and Business Policy, Pearson Education.
- [9] Olszewskd, R., & O'Hanlon, T. (2001). Help Benchmark CMMS software. Maintenance technology.
- [10] Andersen, B., & Pettersen, P. (1996). Benchmarking handbook: Step-by step instruction. Chapman; London, UK.
- [11] Haroun A. E. & Duffuaa S. O. (2009). "Maintenance organisation", in *Handbook of Maintenance Management and Engineering*. London: Springer.
- [12] Márquez A. C., (2007). The maintenance management framework: models and methods for complex systems maintenance. Springer London, UK.

Urban Road Transports E-maintenance supported in HMM

Antón Simões

Dept of Mechanical
Engineering
Instituto Superior de
Engenharia de Coimbra
Rua Pedro Nunes – Quinta da
Nora, 3030-199 COIMBRA
PORTUGAL

assimoes@isec.pt

José Torres Farinha

Dept of Mechanical
Engineering
Instituto Superior de
Engenharia de Coimbra
Rua Pedro Nunes – Quinta
da Nora, 3030-199
COIMBRA
PORTUGAL

torresfarinha@
torresfarinha.com

Inácio Fonseca

Dept of Electrical
Engineering
Instituto Superior de
Engenharia de Coimbra
Rua Pedro Nunes – Quinta
da Nora, 3030-199
COIMBRA
PORTUGAL

inacio@isec.pt

Fernando Maciel

Barbosa
INESC Tec
Instituto Superior de
Engenharia de Coimbra
Faculdade de Engenharia,
Universidade do Porto
Rua Pedro Nunes – Quinta
da Nora, 3030-199
COIMBRA
PORTUGAL

fmb@fe.up.pt

ABSTRACT

The e-maintenance is expanding in many industrial domains. Urban transports road sector is being revealed as one of the most important in its application. E-maintenance is synonymous of effectiveness and efficiency, based on most recent technological resources. It helps to anticipate diagnosis and prognosis conducting to a better organizations performance. This is the way to increment data precision and the decision confidence level. In summary, e-maintenance is a power tool to help understanding sudden changes in the vehicles states, reducing the human intervention. It integrates an automatic logistics supported in the use of Information and Communication Technology (ICT). The project described in this paper is being carried out using advanced intelligent remote sensing systems. This is a way to fight against high emission from effluents of vehicles and, at the same time, to reduce costs in companies including municipal entities. The project can be extended to alert for the inspection needs of the vehicles that circulate in a certain street arch. This is also a way to contribute reducing the environmental impact and to improve the quality life in urban centres.

They are used several tools, equipment and powerful and specific algorithms designed in *MatLab* software. Some equipment integrates a remote sensing device, as a digital high resolution camera, emission gas analyser, opacimeter and a sonometer. The diagnosis, the prognosis and the eventual alarm to initiate an intervention is made in useful time avoiding fault occurrence and enormous damages. There are many types of indicators that can be used to measure equipment condition, as is the case of several types of pollutant emissions such as NOx, COx, HC, PM and noise. The problem is to construct a prediction model that conjugates all these variables and following their evolution along and accurately predicting of present and next states. This type of problems can be solved through a Hid-den Markov Model (HMM) that has shown to have the adequate characteristics to manage the complexity associated to many different variables, taking into account the specificity of this equipment type. It is always present the alignment of the maintenance process with the business and operational process to achieve organisational and political objectives.

Keywords

e-maintenance, ICT, Urban, Road Transports, State, Pollutant Emissions, Diagnosis, Prognosis, HMM.

1. INTRODUCTION

One of the most usual strategies for urban road transports maintenance is the periodic or regular time intervention. Other methodology, usually used in association with this option, is the on-condition maintenance, using as condition variable the oil properties, with the objective of extending engine life cycle.

E-maintenance is based on several views and perspectives. It is considered strictly a complementary tool to e-manufacturing. One possible definition for e-maintenance is: Maintenance system is an architecture which includes the resources, logistics and management necessary to enable proactive decision process execution. This system includes e-technologies (i.e. ICT, wireless, web-based, sensors technologies and digital cameras) but also, e-maintenance activities (logistics and processes) such as e-monitoring, e-diagnosis, e-prognosis, and so on. The e-maintenance enables four main different maintenance facets such as:

- Remote maintenance
- Predictive maintenance
- Real-time maintenance
- Collaborative maintenance

For buses, this innovative paradigm comprises soaked intelligent equipment, making them capable of transmitting real-time data such as alarms and events based on their current condition. These warnings, alarms or events, as a part of the e-maintenance solution are used in evaluation of condition based maintenance (CBM).

Maintenance-related information can be used to increase the efficiency and effectiveness of the transport services and system, but also to provide the consumer of the transport with informative internet services. For example, information about unexpected maintenance activities within a transport system, timetables, delays, maintenance process to enable opportunistic maintenance in order to reduce the negative impact (Biswal and Parida 2010)).

Simultaneously, the same information can be correlated to spatial data in order to provide better decision-support for a route planning service aimed at the consumer of the transport service. Hence, the provision of information services can be considered as essential enabling services that complement the transport services and contribute to increased satisfaction of the service consumer. Technicians and managers are known the important role of oil analysis applications in fleet predictive maintenance (Monnin, 2011). This significantly reduces the reaction time to critical fleet maintenance issues and maximizes overall equipment and vehicle reliability.

The proposed model (Figure 1) characterizes the integrated platform. Is denominated Buses Fleet Ecological e-Maintenance (BFEM), and contains the following subsystem objects:

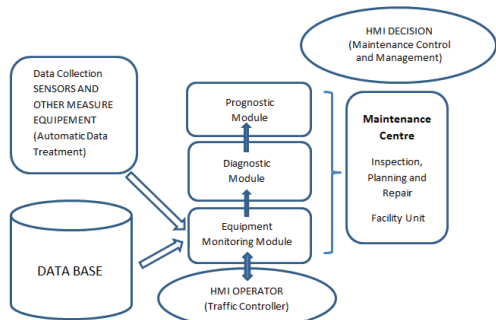


Figure 1. Buses Fleet Ecological e-Maintenance

The model uses a sub-model denominated ecological predictive maintenance (EPM) with environmental indicators. In ecological predictive maintenance (EPM) the use of emissions spectrum together with the HMM is the innovative procedure here presented. The motivation for this is the usefulness of emissions spectrum and the rationality behind it which can without problems be understood by the specialists carrying out condition monitoring with the aid of wireless systems.

2. ECOLOGICAL MAINTENANCE PERFORMANCE INDICATORS

In order to use the measured activity and condition monitoring signals in an ideal way to support the diagnosis and prognosis phases of e-Maintenance refined signal analysis should be used.

Ecological Maintenance Performance Indicators (EMPI) are associated with the pollution mitigation, health, safety, reduction of down-time, costs and wastes, and the enhancement of capacity utilization, productivity and quality. Therefore, EMPI evaluate the actual conditions with a specific set of statistics and reference conditions (requirements / targets).

An e-maintenance framework that is proposed for maintenance management is shown in Fig. 2. The local platform consists of the machinery and systems to be monitored and the condition monitoring system themselves. The data pertaining to the performance and health of the machinery and systems are recorded by the condition monitoring system (Verma, 2010).

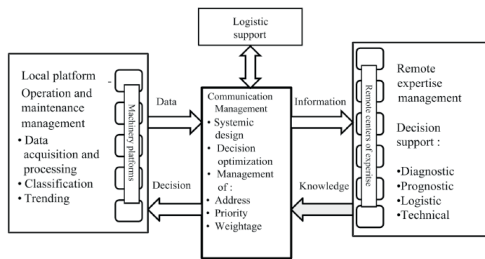


Figure 2. Framework for e-maintenance (Verma, 2010)

The processing power of processors is very important and some of the sensors already have this ability, providing information for this typology of analyses. In the light of reviews it seems that, even though quite a lot of research has taken place and numerous papers about several signal analysis techniques have been published, no essential methods have gained viable success. To perform the acquisition of the monitored signals an acquisition system is proposed in Figure 3. This system is composed by a Linux Server running a MySQL Database System to save signal data. For general implementation and interface with transport systems, four solutions for signal acquisition are presented:

- Ethernet PLC acquisition system based on Beckhoff technology;
- Low cost ARM microcontroller system with Ethernet connectivity;
- CompactRIO System for LabView rapid research development;
- Compact PC board for local installation.

The acquisition signals are acquired on each solution, packed and sent via Ethernet to the MySQL Database.

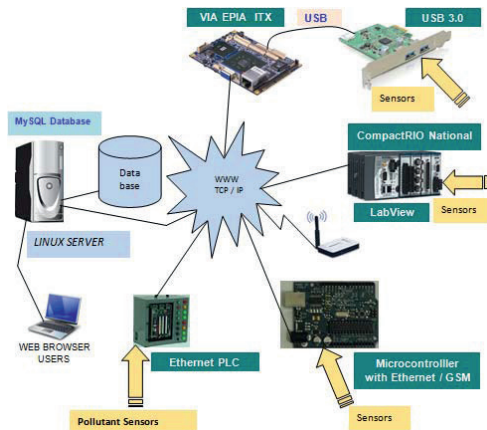


Figure 3. Data Acquisition System

Typical key performance indicators used in fleet management includes operating cost, asset availability, lost time injuries, number of passengers. In our analyses, we defined an emission spectrum to characterize the pollution impact and to infer the

vehicle class state (Fig.4). This matrix includes effluents and noise. At the same time is used the Vehicle Specific Power (VSP).

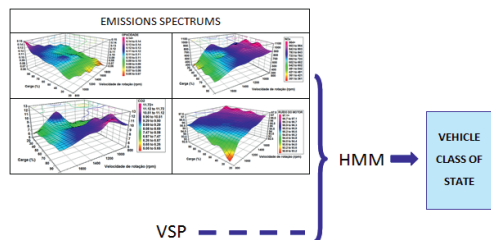


Figure 4. Vehicle class of state evaluation

Emissions spectrums (ES) and overall equipment effectiveness (OEE) are used as a key performance indicator for the fleet maintenance needs in its continuous search for new ways to improve the environment and reduce the downtime, costs and waste, to operate more efficiently, and to achieve greater capacity.

The four elements of ES (Opacity (or PM), NOx, CO2 and Noise), maintenance and service quality, OEE, and availability, help to determine the impact in the environmental and the performance of an individual bus.

3. ECOLOGICAL HEALTH MONITORING

Preventive maintenance and scheduled service activities are triggered by sensors mounted on side of the roads or by vehicle meter readings. All of the e-Maintenance features are combined to help and simplify the fleet maintenance management process. Every time and anywhere platform provides fleet managers with the reporting tools they need to oversee all aspects of the operation, while improving satisfaction for everyone involved in the fleet management process. In development of Buses Fleet Ecological e-Maintenance (BFEEM), an e-Platform for Managing Bus Maintenance Operations, are used the following resources / modules:

- **Technical Management:** Tools and methods of condition monitoring, state diagnoses and prognostic.
- **Data Collection and Reporting:** Use of the latest information technologies and the concepts of management e-maintenance, internet wireless, mobile connection and protocols to collect data and develop dysfunction reports.
- **Modeling and Technical Data Analysis:** Tools that support advanced levels that permits to obtain HMM analysis, environmental impact, cause-consequence analysis, maintenance-engineering analysis, reliability block diagram, FMECA, FTA, ETA, life-cycle cost, tolerance analysis and part count analysis. Model matrix and response matrix with predicted values. Uses MATLAB as support.
- **System Design and Logistics:** Solutions capable to streamline data and information among the departments and institutions, in view to aid business and management decisions. Besides to assure the design of maintenance echelons and selection of variables, factors and levels, evaluate redundancy and diversity, modularity and

diagnostics, reliability versus maintainability, tradeoff studies, part control program, RAM procurement specifications, preventive maintenance program versus ecological predictive maintenance and spare part program.

The relationship between symptoms and dysfunctions play an important role in order to make available a certain understanding (i.e diagnosis) of the corresponding health condition. Many steps are been implemented that to permits anticipate that the ordering of spare parts and their management will in fact become almost always automatic. Nevertheless, to define the classes it becomes necessary to adopt a set of limits, international standards, environmental specifications or any other norms, according to each specific situation. The BFEEM architecture is shown in Fig. 5.

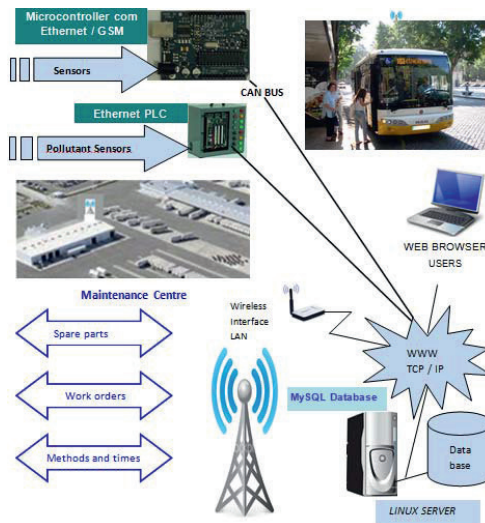


Figure 5. Bus fleet e-maintenance Lay-out

The usage of emissions spectrum data as the singular evidence for diagnosis shows that exist a strong correlation with engine states. The reality is that automatic diagnosis has not turned out to be widespread even though a lot of research has developed lots of applications.

In the future when data becomes obtainable from the increasing number of sensors the signal analysis has to be automated so that it then supports programmed diagnosis and prognosis. One enormous challenge is the blend of emissions signals together with dynamic vehicle signs to determine the state physical and ecological of fleets. When emissions spectrum and traffic signals are automatically collected and the accessible technologies processes them with advanced and refined procedures which would not appear straightforwardly understandable for humans are achieved new steps in maintenance management. This can lead to sort of black box solutions i.e. the controller person depends on the routines informatics and technologies not perceptive for the users. In the forthcoming years we can vaticinate fairly a lot of study in this area.

It is astonishing how limited the use of automatic diagnosis still is in transports. The prove of this is a very scarce reliable systems are used to support condition monitoring even though a lot of research has been tried to resolve the problem. Is still nebulous the state of affairs, although the article point out new ways. The major challenge is really to adequate the public infrastructures in order to be able to mount sensors, signals analysis tools and communication nets, in order to be able to produce something that is truly useful. Still the usage of correct sensors with proper signal analysis techniques constitute the challenge, there is no change in operators mentality.

On the other and, the new scientific developments point out that in the following years the enriched understanding of wear of the components of the machinery and the creation of new informatics algorithms will help in the automation of diagnosis. It can be expected that in the future supplementary process data will be available and also simulation models will be available to provide added information to support monitoring, diagnosis and prognosis. Up-to-date e-Maintenance solutions could even now carry out all the ordering and work force management automatically. In scenario of things going wrong is important the role of neural networks. To manage different fleets is important to know the service context, operational context and performance context (Monnin, 2011).

The significant health indicator can be defined according to the corresponding component and context. For instance, abnormal behavior can be caused by the system environment. In that case the contextual information do not only concern technical or service context level. To evaluate the health condition situation of component can be used different criteria or context. From the diagnosis point view, abnormal behaviors, which are depicted through the health condition, can be defined by symptom indicators. Finally, the performance context is linked to the key purpose of the fleet and defines, to some extent, the focus of optimization (example focused on availability or costs). The relationship between symptoms and dysfunctions is also considered in order to make available a certain understanding of the corresponding health condition. This task is normally named diagnosis (Figure 6).

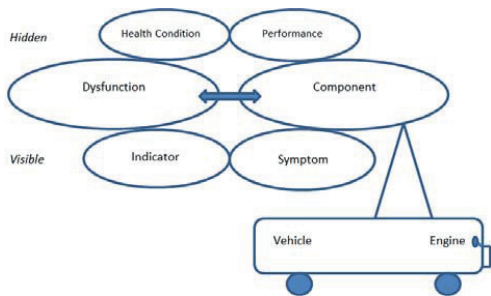


Figure 6. Fleet model conceptualization

E-Platform for Managing Bus Maintenance Operations is a program for the systematic monitoring and evaluation of the various aspects of technical activities and services, to ensure that the fleet is expected to meet, or continues to meet, established availability performance goals and to make sure that demands is

according to enterprise plans. It is based on the fact that any services should be suitable for the intended purpose. It is a set of activities intended to ensure that established availability performance requirements are met and pollution emissions are been mitigate (Fig. 7). In the operation phase, observations of the Fleet Availability performance should be used to evaluate the need for corrective actions, i.e. modifications, and improvements. The necessary feedback should identify (Ahmadi, Karim et al. 2010):

- i) How well the fleet is performing in comparison to established goals and requirements.
- ii) Where and why the deviations occur, if the performance goals and requirements have not been achieved.
- iii) How to define cost-effective solutions to achieve the necessary improvements.

1-Availability factors		
1.1-Reliability Design Tolerances Quality control Operating Conditions	1.2-Maintainability Organization Resources Special tools Spares Accessibility Modularization	1.3-Maintenance support and configuration flexibility Maintenance program Equipment and facility Spare parts and tools Maintenance crew Training Subcontracting Exploitation profile
2-Production availability Operation availability or fleet availability Capacity Mission and demand Buffer time Fleet rotation Consequences of failure Certifications Insurance		3-Service Availability Compensation Substitution of service

Figure 7. Availability factors groups

The main aspects reported in this paper are the following ones: why to adopt HMM's; how the model is constructed; how the observed values are read; how the model is trained; why the model is adequate and how to evaluate the perplexity index [1] [2] [3].

4. HMM

The Markov process models correspond to a class of probability models used to study the evolution of a system over time. Transition probabilities are used to identify how a system evolves from one time period to the next. A Markov chain tries to characterize the system behaviour over time, as described by the transition probabilities matrix and the probability of the system initial state.

To be represented by a HMM a process must accomplish the following requisites:

- To be a stochastic process;
- The transition probability from the present state to any other one is independent of how the process reached the current state. So, for computation purposes, the only

necessary state is the previous one; past states are not needed.

A Hidden Markov Model is an extension of a First Order Markov Model:

- The “true” states are not directly observable (are Hidden);
- The observable variables (emissions, in this paper) are probabilistic functions of the hidden states;
- The hidden states behave as a First Order Markov.

A Hidden Markov Model consists of two sets of correlated states (states and emissions in this paper) and the following set of probabilities:

- The hidden states - the “true” states of a system that may be described by a Markov process;
- The observable variables (emissions) - these are the indicators, the “face” of the process; they are “visible”;
- The initial probabilities - for hidden states;
- The transition probabilities matrix - for hidden states;
- The model output symbol probability matrix - where each element represents the probability of outputting a symbol of emissions, taking into account that the model is in a specified state: it’s the observation probabilities matrix.

This paper shows that on-condition maintenance prediction can be based on a Hidden Markov Model (HMM) as demonstrated along next sections, that also reveal the model performance and simplicity. The model integrates an initial state probability vector whose elements represent the probabilities of the system to start in different classes that correspond to different states. In fact, the emission of any Diesel Engine vehicle is a random process, depending of many variables such as the driver behaviour, exploitation conditions, temperature and so on.

Usually, the system is planned to make cycles between interventions, beginning at the zero instant (usually when the vehicle is new) and evolving along time with the various maintenance interventions and states. Frequently it would be of interest using HMMs associated with continuous observation densities. However and in order to allow this, some restrictions have to be applied to the model probability density function in order to ensure that the parameters of such function can be reestimated in a consistent way. A continuous HMM, as a discrete one, has the following properties:

- The residence time in a state is Markovian (memoryless process);
- The choice of the next state is affected at the instant of transition and depends on the current state only, not on the past states: so, it is Markovian.

However, the following two reasons forbid the validation of the conditions of a continuous HMM:

- 1) The process stay in each state doesn’t necessarily follow an exponential distribution;
- 2) In the context of this paper it depends on acceleration, weather and many variables that continuously change. Additionally, by practical reasons, it is not feasible to continuously measure emissions; only sampling is possible.

One greater difficulty is data classification. Researchers are very inexperienced when they assume that a sophisticated classifier can resolve lots of problem. Constitute the basic evidence that a good classifier means a better performance in sense to avoid crack the system. In this research was used the conventional statistics. However the use of neural networks can to be an alternative.

5. EMISSIONS AND HMM STATES

The case of Diesel emissions assumes the following specific situation: there is an initial state (usually when the equipment is new), the state at instant zero. Next states evolve until the limits imposed by international standards, environmental rules and some requisites of each real situation are reached. Figure 8 shows a generic evolution for the Diesel emissions context (Simões, 2011).

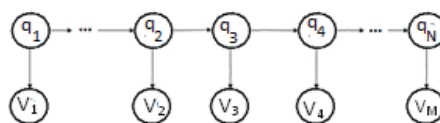


Figure 8. Hidden states and observed symbols in a Diesel engine

- $Q : \{q_1...q_N\}$ - Set of values for the hidden states - state vectors / states library;
- $V : \{V_1...V_M\}$ - Set of values for the observations - observation vectors / emissions library.

Figure 9 shows the temporal sequence of each vector associated to the transition matrix and emission matrix, as follows:

- $A = \{a_{ij}\}$ - State transition probabilities matrix;
- $B = \{b_{ik}\}$ - State observation probabilities matrix.

To start the model it is necessary to know the vector state for instant zero:

- $\Pi = \{\pi_i\}$ - Initial state probability matrix.

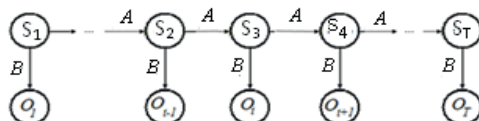


Figure 9. HMM with transition probabilities matrices

So, the HMM model is given by: $\{Q, V, \Pi, A, B\}$ (1)

The HMM implementation is supported by the following steps:

1. Compute the probability of the observation sequence;
2. Based on the observation sequence, compute the most likely hidden state sequence;
3. Based on the observation sequence and set of possible models, answer the question: “which is the model that most closely fits the data?”.

Given an observation sequence and a model, the probability of the observation sequence is given by:

$$O = (o_1 \dots o_T), \lambda = (A, B, \Pi) \quad (2)$$

Then, it is necessary to compute $P(O|\lambda)$. One way to do this is to consider that the probability of an observation sequence is the sum of the probabilities of all possible state sequences conducting to that state. However:

- Naïve computation is very expensive as, given T observations and N states, there are N^T possible state sequences;
- Even for small HMMs, e.g. $T=10$ and $N=10$, there are 10 billion different paths.

A possible solution for this problem consists of using dynamic programming. This can be accomplished by the following steps

$$P(O|S, \lambda) = b_{s_1 o_1} b_{s_2 o_2} \dots b_{s_T o_T} \quad (3A)$$

$$P(S|\lambda) = \pi_{s_1} a_{s_1 s_2} a_{s_2 s_3} \dots a_{s_{T-1} s_T} \quad (3B)$$

$$P(O|\lambda) = \sum_S P(O|S, \lambda) P(S|\lambda) \quad (3C)$$

$$P(O, S|\lambda) = P(O|S, \lambda) P(S|\lambda) \quad (3D)$$

that can be resumed by:

$$P(O|\lambda) = \sum_{\{S_1 \dots S_T\}} \pi_{s_1} b_{s_1 o_1} \prod_{t=1}^{T-1} a_{s_t s_{t+1}} b_{s_{t+1} o_{t+1}} \quad (4)$$

that is, the forward-backward procedure, as shown in next section.

The HMM has several unsolved problems that conduce to several solutions, according to the following procedures:

- I. Evaluation:
 - Problem - There is a set of possible HMMs and a given number of observations in which HMM more probably generated the given sequence?
 - Solution
 - ✓ Compute the probability of the observed sequences for each HMM;
 - ✓ Choose the one with the highest probability;
 - ✓ Use the Forward algorithm to reduce complexity.
- II. Decoding:
 - Problem - Given a particular HMM and an observation sequence, which is the most likely sequence of underlying hidden states that might have generated this sequence of observations?
 - Solution:
 - ✓ Compute the probability of the observed sequences for each possible sequence of underlying hidden states;
 - ✓ Choose the one with the highest probability;
 - ✓ Use the *Viterbi* algorithm to reduce complexity.
- III. Learning:
 - Problem - Estimate the probabilities of the HMM from the training data?
 - Solution:

- ✓ Train with labelled data:
 - Transition probability $P(q_i, q_j)$ = (number of transitions from q_i to q_j) / (total number of transitions of q_i);
 - Observation probabilities matrix $P(q_i, V)$ = (number of symbol V occurrences in state q_i) / (number of all symbol occurrences in state q_i);
- ✓ Train with unlabeled data:
 - *Baum-Welch* algorithm basic idea:
 - Maximize the probability of the observation sequence, given the model.
 - Estimate new probability from the previous HMM until $P(\text{current HMM}) - P(\text{previous HMM}) < \epsilon$ (a small number).

In what concerns to labelled and unlabelled data sections [4] and [5] deal with this problem.

6. FORWARD-BACKWARD ALGORITHM

The prediction of hidden states is an extremely complex problem, what implies the implementation of an algorithm to more efficiently solve it [6] [7]. A possible and efficient solution consists of using dynamic programming. The first steps are:

- Intuition – that is, the probability of the first t observations be the same for all possible $t+1$ length state sequences. This can be defined by eq. 5

$$\alpha_t(i) = P(O_1, O_2, O_3, \dots, O_t, S_t = q_i | \lambda) \quad (5)$$

that is the basis of the forward procedure.

The calculation of $\alpha_t(i)$ is achieved by the following steps:

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N \quad (6A)$$

$$\alpha_{t+1}(j) = P(O_1, \dots, O_t, O_{t+1}, S_{t+1} = q_j | \lambda) =$$

$$\sum_{i=1}^N P(O_1, \dots, O_t, O_{t+1}, S_{t+1} = q_j | S_t = q_i, \lambda) P(S_t = q_i | \lambda) =$$

$$\sum_{i=1}^N P(O_1, O_2, \dots, O_t | S_t = q_i, \lambda) P(S_t = q_i | \lambda) P(O_{t+1}, S_{t+1} = q_j | S_t = q_i, \lambda) =$$

$$\sum_{i=1}^N P(O_1, \dots, O_t, S_t = q_i | \lambda) P(O_{t+1} | S_{t+1} = q_j, S_t = q_i, \lambda) P(S_{t+1} = q_j | S_t = q_i, \lambda)$$

$$\alpha_{t+1}(j) = b_j(O_{t+1}) \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] \quad (6B)$$

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}), \quad 1 \leq t \leq T-1 \quad e \quad 1 \leq j \leq N \quad (6C)$$

Next step is the *backward* procedure, which allows the computation of the observations probability sequence given a known state at instant t . This can be done according to the following equations:

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, O_{t+3}, \dots, O_T | S_t = q_i, \lambda) \quad (7)$$

- Initialization:

$$\beta_T(i) = 1, \quad 1 \leq i \leq N \quad (8)$$

(TCRP-SINTHESIS-81)

- Induction:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad t = T-1, T-2, \dots, 1 \quad e \quad 1 \leq i \leq N \quad (9)$$

The final steps for decoding the solution arc

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (10)$$

that is, the forward procedure, and

$$P(O|\lambda) = \sum_{i=1}^N \pi_i b_i(O_1) \beta_1(i) \quad (11)$$

that is, the backward procedure.

7. Viterbi algorithm

Now we'll try to find the state sequence that best explains the observations. This can be done by the Viterbi algorithm, synthesized by equation 12.

$$S^* = \arg \max_S P(S, O|\lambda) \quad (12)$$

To evaluate this equation, an auxiliary variable $\delta_t(i)$ is defined. This variable corresponds to the maximum result (or higher probability) of the recorded sequence of observations, assuming that the end state is stage q_i :

$$\delta_t(i) = \max_{S_1, S_2, S_3, \dots, S_{t-1}} P[S_1 S_2 S_3 \dots S_{t-1}, S_t = q_i, O_1 O_2 O_3 \dots O_t | \lambda] \quad (13)$$

The algorithm may be compactly stated by:

- Initialization:

$$\delta_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N \quad (14)$$

$$\psi_1(i) = 0$$

- Recursive computation:

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(O_t), \quad 2 \leq t \leq T \quad e \quad 1 \leq j \leq N \quad (15A)$$

$$\psi_t(j) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad 2 \leq t \leq T \quad e \quad 1 \leq j \leq N \quad (15B)$$

- Termination:

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)] \quad (16A)$$

$$S_T^* = \arg \max_{1 \leq i \leq N} [\delta_T(i)] \quad (16B)$$

Next step consists of computing the most likely state sequence by backtracking:

$$S_t^* = \arg \max_{1 \leq i \leq N} [\delta_t(i) a_{i S_{t+1}^*}] \quad (17A)$$

$$S_t^* = \psi_{t+1}(S_{t+1}^*), \quad t = T-1, T-2, \dots, 1 \quad (17B)$$

The result is the generation of the sequence:

$$S_1, S_2, S_3, \dots, S_{t-1}, S_t \quad (18)$$

8. BAUM-WELCH ALGORITHM

The final step to completely define the HMM is parameter calibration: given an observation sequence, which is the model that most likely produces that sequence? Given a model and observation sequence, update the model parameters that best fit the observations.

$\xi_t(i, j)$ is defined as the probability of the system to be in state q_i at time t and in state q_j at time $t+1$

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \quad (19)$$

Eq. 19 gives the probability of traversing an arc.

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad t = 1 \dots T \quad (20)$$

Eq.20 gives the probability of being in state i at time t . Now it's possible to compute new estimates for the model parameters:

$$\hat{\pi}_i = \gamma_1(i) \quad (21)$$

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^T \gamma_t(i)} \quad \hat{b}_j(k) = \frac{\sum_{t=1}^T \gamma_t(j)}{\sum_{t=1}^T \text{For } O_t = V_k \gamma_t(j)}$$

9. PERPLEXITY MEASUREMENT

The performance of the HMM model can be measured in two ways:

- 1) by classification accuracy;
- 2) by perplexity.

Classification accuracy is the amount of correct predictions of hidden states vector divided by the total amount of hidden states under prognostication. However, for relatively small data sets, the classification accuracy is a noisy measure since each sample can be assigned to only one class. Therefore, a better measure is the perplexity of the test data set, which measures the confidence of the predictions of the classifier that is defined as a function of the average of log-likelihoods L of the N data sequences, formally denoted by

$$L_t = \log p(q_i | O_{1..T_s}^i, \lambda) \quad \text{perp} = e^{-\frac{1}{N_s} \sum_{i=1}^{N_s} L_t} \quad (22)$$

where $O_{1..T_s}^i$ stands for the i th sequence of observations of length T_s , and q_i is the type of hidden states i ; N_s is the number of sequences and λ the model parameters. The best possible perplexity is 1, where the correct task type is predicted with a probability 1. Additionally, a perplexity of 3 corresponds to random guessing with a probability of $1/3$ for each of the hidden

states [8].

10. IMPLEMENTATION

Maintenance performance indicators (MPI) are linked to the reduction of down-time, costs and wastes, and the enhancement of capacity utilization, productivity, quality, health, safety and environmental impact, identifying if the fleet are operating more or less efficiently. Therefore, MPIs compare the actual conditions with a specific set of reference conditions (requirements/targets).

The three elements of OEE, availability, performance speed and quality, help to determine the impact of the performance of an individual piece of equipment. Automated fleet management employed the concept of e-maintenance in transport enterprises corresponds to a system that provides: instant email notification of dysfunctions, performance issues or faults, often before the user would even know there's a problems, automatic monitoring of maintenance requirements and performance, as well as consumables management (for example, receive a notification when parts need substitution). The program enables the Technical Center or a Service Centre to remotely access a Bus Central Maintenance Computer to troubleshoot.

The subject of this paper is the on-condition maintenance of Diesel engines, that can be managed, according to the above exposed, by a state matrix with the following variables: {NOx, COx, HC, PM, NOISE}, where the various combinations / interceptions represent the emission observable variables - {V1, V2, V3, V4..... VM}. According to this, there are five possible states including the initial one. The thresholds used for states segmentation are 1-Excellent; 2-Good; 3-Tolerable; 4-Dysfunctional; 5-Bad. The classes of states are identified by q1, q2, q3, q4, q5. Consequently, the transition matrix has a dimension of 5x5. The number of decision thresholds for each ecological variable is four: 1-Excellent; 2-Good; 3-Acceptable; 4-Poor or bad. The current model incorporates only three emission variables {NOx, PM, and NOISE}.

The real and hidden first states can overlap if the equipment is assumed as new or just after a maintenance operation. In general:

$$\pi_i = P[S_i = q_i], \quad 1 \leq i \leq N \tag{23}$$

where π_i = Number of times in state q_i at time $t = 1$.

Now, each element of matrix q must be computed. The procedure is:

$$a_{ij} = P[S_{t+1} = q_j | S_t = q_i] = \frac{\text{number of transitions from } q_i \text{ to } q_j}{\text{number of times in state } q_i} \tag{24A}$$

$$b_j(k) = \frac{\text{nr of times in state } q_j \text{ with the symbol } v_k}{\text{number of times in state } q_j} \tag{24B}$$

The four levels associated to each of them presuppose the existence of 64 combinations. For simplicity, the 64 combinations are grouped into 11 classes of emission only, designated by V1, V2, V3, V4,V11.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \tag{25}$$

The library of emissions is much more complex since it integrates the environmental impact of different variables:

$$B = \begin{bmatrix} h_1(1) & h_1(2) & h_1(3) & h_1(4) & h_1(5) & h_1(6) & h_1(7) & h_1(8) & h_1(9) \\ h_2(1) & h_2(2) & h_2(3) & h_2(4) & h_2(5) & h_2(6) & h_2(7) & h_2(8) & h_2(9) \\ h_3(1) & h_3(2) & h_3(3) & h_3(4) & h_3(5) & h_3(6) & h_3(7) & h_3(8) & h_3(9) \\ h_4(1) & h_4(2) & h_4(3) & h_4(4) & h_4(5) & h_4(6) & h_4(7) & h_4(8) & h_4(9) \\ h_5(1) & h_5(2) & h_5(3) & h_5(4) & h_5(5) & h_5(6) & h_5(7) & h_5(8) & h_5(9) \end{bmatrix} \text{ (NORMA_NP_EN_5 0126)}$$

The challenge is not just to diagnose the state associated to a combination of emissions but also to prognoses the next emission level and correspondent states of health.

Table 1 – Main Outputs of Ecological HMM

Emissions at time	t=1	t=2	t=3	t=4	t=5	t=6	t=7
Classification of emissions	V ₄	V ₅	V ₇	V ₇	V ₄	V ₉	V ₁₀
Real hidden state	S ₁ = q ₁	S ₂ = q ₄	S ₃ = q ₄	S ₄ = q ₅	S ₅ = q ₂	S ₆ = q ₅	S ₇ = q ₅
Hidden state classification	q ₁	q ₄	q ₅	q ₅	q ₂	q ₅	q ₅

The updating of the initial state probability matrix has been carried out based on models of degradation taken from operating conditions. From a predictive diagnosis point of view existing alarm monitoring systems are sufficient once they allow failure to be anticipated. Once the alarm occurs, the remaining time to failure is short for preventing it. Moreover, the cause identification of such alarms must be analysed subsequently. The knowledge-based model proposed allows providing such monitoring data with the corresponding context at different levels.

Figure 8 presents an example of fleet-wide expertise retrieval results. For the given engines of the fleet (Table 1), some diagnostic results are proposed and summarized. With such a system, the expert, in face with a particular situation, can make any association to find out the closest cases with the case to solve and shall concentrate on the most frequent degradation modes already observed. From the different contextual information available, the system helps understanding the behaviour without hiding its complexity with too simplistic rules.

11. CONCLUSIONS

The point of discussion in this paper is: to present opportunities and challenges regarding e-maintenance.

Based on the multiple researches we tried to promote some of the opportunities and challenges as mentioned in the purpose of the paper. We have made an analysis of these opportunities and

challenges and identified two main areas: technology and business. This paper will serve as the basis for more in-depth study based on extensive literature review and empirical findings. In future work, a study can be done on how e-maintenance can influence business model of an organization and the environmental improvement in the cities.

Fleet-wide PHM requires knowledge-based system that is able to handle contextual information. Diagnosis and maintenance decision making processes are improved by means of semantic modeling that deals with concepts definition and description. In this paper, a knowledge model is proposed. The application context allows consider fleet component similarities and heterogeneities. Data of the monitored bus are considered within their context and enhance the identification of the corresponding health condition.

This paper gives an overview of an open systems approach to e-maintenance with an innovative proposal of how a conventional enterprise can be transformed to a fully automated e-maintenance solution. The weaknesses of the existing transport operators are known. To implement new solutions is very important, in view to optimize organizational profits. It also mentions the capabilities of the proposed architecture and addresses certain challenges faced in order to enable an open framework. The detailed cost benefit analysis for the framework is beyond the scope of this paper and will be discussed in a follow up paper.

The findings so far suggest that the different traditional practices used in preventive and condition based maintenance strategies require building customised solutions. An open system solution can tackle the associated problem in fairly cost-effective manner. A new development, ubiquitous and intelligent system of e-maintenance using standards "on demand" shall facilitate interoperability with existing legacy systems. A prospect of this work will be to generate a virtual framework of the system to do performance benchmarking of the integrated system based on real data gathered through the real system. This shall further help in evaluating whether or not the solution supports the transport sector wide best practices and the total maintenance process.

The maintenance of Diesel engines and the buses through on-condition maintenance planning allows the improvement of the intervals between interventions, based on the following condition variables: NOx, PM and NOISE. To plan maintenance with such variables, a Hidden Markov Model (HMM) was developed. The model showed to be adequate to manage the complexity associated to these variables. The model uses the Baum-Welch and Viterbi algorithms. The efficiency of prediction of the model depends on the frequency of data collection and the characteristics of systems for automatic detection of environmental impacts. This is a new paradigm towards a better world with impacts on emissions mitigation, especially in urban areas. The success of the model is the versatility of its application in all fields of evaluation of health status. Future work will consist in the development of a decision analysis tool for evaluating the vehicles maintenance level and needs, designed for urban transports applications. To measure the emissions, the system will use remote sensing devices.

12. REFERENCES:

- [1] Ahmadi, A. et al. (2010). *Prerequisites for a Business-oriented Fleet Availability Assurance Program in Aviation*. 1st international workshop and congress on eMaintenance 2010.
- [2] Biswal, M. et al. (2010). *An integrated approach for open e-Maintenance: Opportunities and challenges*. 1st international workshop and congress on eMaintenance 2010.
- [3] Bilmes, J. (2002). *What HMMs Can Do*. UWEE Technical Report Number WEETR-2002-0003; Department of Electrical Engineering; University of Washington; Box 352500 Seattle, Washington 98195-2500 bilmes@ee.washington.edu.
- [4] Clarkand, N. et al. (2003): *Predictive Tool for Emissions from Heavy-Duty Diesel Vehicles*. Department of Mechanical and Aerospace Engineering, Argonne National Laboratory, Building 362.
- [5] Cunha, A. (2003): *Métodos Probabilísticos para Reconhecimento de Voz*. Technical Report TR-03-04, Instituto de Matemática Pura e Aplicada, Laboratório VISGRAF- Brasil.
- [6] Frey, C. et al. (2007). *Comparing Real-World Fuel Consumption for Diesel and Hydrogen Fueled Transit Buses and Implication for Emissions*. Elsevier Science.
- [7] Monnin, M. et al. (2011). *Fleet-wide health management architecture*. Annual Conference of the Prognostics and Health Management Society, 2011.
- [8] Monnin, M. et al. (2012). *Fleet historical cases for predictive maintenance*.
- [9] Rabiner, L. et al. (1983): *On the Application of Vector Quantization and Hidden Markov Models to Speaker-Independent, Isolated Word Recognition*. The Bell System Technical Journal.
- [10] Rabiner, L. (1989): "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition". Proceedings of the IEEE, 77(2).
- [11] Seeger, M. (2002): *Learning with labeled and unlabeled data*. Institute for Adaptive and Neural Computation. University of Edinburgh. 5 Forrest Hill, Edinburgh EH1 2QL. seeger@dai.ed.ac.uk. December 19, 2002.
- [12] Simões, A. et al. (2008): *Ecological predictive maintenance in hospital equipment*. Proceedings of 20th Congress of the International Federation of Hospital Engineering (IFHE), Barcelona, 19-22 October.
- [13] Simões, A. et al. (2009): *Buses Degradation Based on Exploration Conditions*. Proceedings of the 3rd WSEAS Int. Conf. on ENERGY PLANNING, ENERGY SAVING, ENVIRONMENTAL EDUCATION. ISSN: 1790-5095; ISBN: 978-960-474-093-2. Pp 141-149.
- [14] Simões, A. et al. (2010): *Ecological predictive maintenance in urban fleets*. Proceedings of The 5th IASME / WSEAS International Conference on ENERGY & ENVIRONMENT (EE'10). ISSN: 1790-5095 384, ISBN: 978-960-474-159-5. Pp 46-53.
- [15] Simões, A. (2011). *Mantenção Condicionada às Emissões Poluentes em Autocarros Urbanos - Diagnóstico por Cadeias Escondidas de Markov*. Doctoral Thesis, Lisbon's High Institute of Technology (IST).
- [16] Simola, J. et al. (2008): *Using hidden Markov model to uncover processing states from eye movements in information search tasks*. Cognitive Systems Research, volume 9, number 4, pages 237-251. Elsevier Science.

- [17] Verma, A. K. et al. (2010). *A Systemic Approach to Integrated E-maintenance of Large Engineering Plants*. International Journal of Automation and Computing, International Journal of Automation and Computing.
- [18] Zhang, J. (2002): *Markov Chain and Hidden Markov Model*. Purdue University.
- [19] Zhou, D. et al. (2005): *Learning from Labeled and Unlabeled Data Using Random Walks*. Max Planck Institute for Biological Cybernetics. Spemannstr. 38, 72076 Tuebingen, Germany.

A new approach for the diagnosis of maintenance state

Hugo Raposo
CEMUC
Coimbra
Portugal
+351 239 790700
hugrap@gmail.com

José Torres Farinha
CEMUC
Coimbra
Portugal
+351 239 790700
torres.farinha@dem.uc.pt

Inácio Fonseca
CEMUC
Coimbra
Portugal
+351 239 790700
Inacio.fonseca@dem.uc.pt

Luís Andrade Ferreira
FEUP
Porto
Portugal
+351 22 508 14 00
lferreir@fe.up.pt

ABSTRACT

This paper presents a holistic model to diagnose the maintenance state in a company regarding the base of the organization, the transversal politics and the new management perspectives.

The diagnosis of the maintenance state allows making the analysis of this company's sector, emphasizing its strong points, its weaknesses and the aspects to improve in the future.

The developed model permits its generalization into any organization of any type of business, what reinforces its importance. The model was conceived to support professionals and managers, with the goal to audit the maintenance departments of each organization and to adopt the procedures that lead to help the reorganization's actions, and the implementation of new techniques and new management models.

This model was validated in the three biggest Portuguese companies of Urban Transportation and the results were analyzed with the respective maintenance managers having corresponding to their empiric perception.

Keywords

Maintenance state diagnosis; Management models; maintenance state

1. INTRODUCTION

Nowadays in the globalized economy, the companies' survival or, in a more general way, the organizations, depends on their innovation ability. For this end, they have been searching unceasingly for new tools that can contribute to their continuous improvement.

However, to get this point, it is important to do a retrospective of the maintenance activity. In fact, as far as we go through reviewing the maintenance history, it is interesting to realize that since the Industrial Age until the 70's, the maintenance function did not evolve much, because the strategies for a continuous improvement did not exist.

Until the beginning of the 70 decade, most part of the industrial plants managed the maintenance in a reactive way, after a stop due to a malfunctioning – this was the corrective maintenance or curative or, in a generalized way, not planned maintenance.

The computers' advent led to the implementation of systematic preventive maintenance strategies in many companies.

By the end of the 70 decade, rises the concept of Reliability Centered Maintenance (RCM), to reduce the increasing volume of Work Orders not planned. The first RCM procedures were very influenced by the safety factors due to its origin in aeronautical industry. At same time, a maintenance philosophy called Total Productive Maintenance (TPM), gained success among Japanese makers. The TPM establishes a partnership between production and maintenance, allowing that basic maintenance operations

(such as cleaning and inspections) would be made by the machines' operators.

In the middle of the 80's, the technological advances in instrumentation and the arrival of the personal computer, it was possible to predict the faults in the equipment, by measuring their condition through vibrations, temperatures and ultra-sound sensors. This technology is often referred as Predictive Maintenance by on-condition controlling.

Other maintenance strategy that has appeared is the Proactive Maintenance that allows extending the equipment functioning cycles through the systematic removal of the failure sources.

With the development of the microcomputers in the 80's, the maintenance area started a process of development, namely in the improving of the information processing and the decreasing of the human dependence with the increasing of equipment availability.

After this long evolutionary process, during which the maintenance area acquired a wide variety of knowledge that came from engineering, informatics, operational research, among others, what permitted to reach a new step of functioning, making pertinent a new and broader activity of maintenance area.

From this perspective, this paper deals with a new tool to help the maintenance management, characterized by to be a holistic model for diagnosis of the maintenance state, in a context that includes the most updated management tools.

The model under discussion try to contribute to the improving of the diagnosis models of the maintenance state, presenting three guidelines: one referring to the organizational perspective; other, to transversal concepts; and the third one, about new maintenance management models. This holistic approach meets the vision presented by [1], on which refers that "with the evolution of the many maintenance concepts and with the development of new approaches and methodologies applied to other management perspectives, namely at quality and production areas, the maintenance activity began to fit and to include those new concepts".

About new diagnosis model of management state, they include, RCM, RBM, FME(C)A, PDCA, SWOT analysis, GUT matrix, JIT, KANBAN, 5S, TPM, Lean Maintenance, RAMS, KAIZEN, SMED, TAUGUCHI, and so on – all of them represent a more added value for the increase of the accomplishment quality of the maintenance activity.

However, in spite of the implementation of these measures, it is essential to understand that the maintenance quality depends on, significantly, from an efficient maintenance management, based on innovative and effective politics that contribute to the increasing of the users' satisfaction.

According to [1], the introduction of integrated management systems in the organizations often becomes not a solution but an additional problem. According to the same author, this is due to the need of reorganization that is required to proceed before the

implementation of the informatics system and a new organic that, necessarily, introduces new variables and, consequently, imposes changes that should be made in a planned way and with the lowest perturbation to the organization.

The reorganization of a maintenance department should be preceded by a diagnosis of its situation and the aspects to correct must be according to the most suitable profile of the organization. The diagnosis, according to [2], on the organizational changes processes, implies the evaluation of the organization actual state, with the goal of determining the needed conditions to reach a new state, more desirable than the actual one.

According to [3], the concept of diagnosis is: "the diagnosis, on the descriptive meaning is a group of prepositions that formulate themselves on the actual state of an organization. It's not a proven conclusion, elaborated from pre-collected information".

One can tell that the methods of management diagnosis are becoming more important for the growing and development of the organizations, being needful to the improving and quality of its products and services. Therefore, the diagnosis is conceived to identify organization activities and exploration of installations that are a target of improvement and efficiency. The auditions of diagnosis have the goal to build a structure that allows the organizations to review, analyze and often recommend improvements on the functioning of the maintenance area and in the productive sector.

The result from that auditing is a work plan that specifies the areas that need improvement, their respective corrective actions and the definition of procedures to the monitoring of the results obtained with the introduction of corrective actions. The main goal of the diagnosis is the identification of the aspects that, after being reviewed, suppressed or improved, bring to an addiction of the company's competitiveness by the increase of its efficacy and effectiveness.

2. HOLISTIC MODEL FOR DIAGNOSIS

To elaborate the diagnosis model for maintenance state presented in this paper, it was taken as the main reference the [4] one, that includes a model that was deeply redesigned, in which were added new options and, consequently, new questionnaires, comprising three perspectives: organization basis; transversal concepts; and new management models. This new approach was designated by Holistic Model for Maintenance Diagnosis State (HMMDS).

Companies, in a general way, haven't given much importance to the maintenance sector and, consequently, to its diagnosis state. However, this activity is becoming more and more important; then it is sought the development of a diagnosis model that rapidly permits obtaining fundamental information about the maintenance state, as well as presenting suggestions for improvement to the weak points. It is also a goal to present a model that can be generalized to any activity sector; this means that it will not be only applicable to small and medium sized companies, but in all companies and also in the most diverse activity fields.

2.1 DIAGNOSIS MODEL

The diagnosis model of the maintenance state is based on a sequence of inquiries on which the evaluation of the answers tells the positioning of the company regarding the state of this department. This audition method was conceived to be a practical instrument of direct utilization by the organizations' responsible technicians in the appreciation of the effectiveness in the use of

their equipments and installations and in the application of systematic improvement measures

The maintenance state diagnosis seeks to support the implementation of an evolutionary step in this activity, placing innovators tools adapted to the state-of-the-art at their disposal. This way is essentially based on two perspectives that guide all its applications:

1. The first one is the introduction of a new and innovative information system, needful tools to the treatment of diverse data regarding the maintenance management as well as the management of the technical assets in general, needed to the organizations that have the goal of reaching an effective management of their life cycle.
2. The second one is the Quality Guarantee Systems and the consequent certification by the adequate regulations in addition to the specific maintenance regulations, which obligate the maintenance function to better interpret their actions and functioning.

The developed model for maintenance diagnosis was implemented in the *Excel* program, what allowed the elaboration of a friendly interface that, simultaneously, helps to automate the model, becoming easier its filling by the enquired and the later data treatment.

To the application of these enquires, the adequacy to any company or activity sector was considered, and there was the preoccupation of do not individualize them, including three perspectives: organization basis; transversal concepts and new management models (table 2.1).

On the organization basis, seven questionnaires were covered: Management of Technical Assets; 1st Level Maintenance; Planning and Safety; Data Base; Maintenance Work; Stocks; Cost Analysis.

Regarding the transversal concepts, the following four questionnaires were covered: RAMS Analysis; RCM; RBM; Transversal Tools.

At last, in the new maintenance management model section, three questionnaires were used: 5S; TPM; LEAN Maintenance.

The final diagnosis is structured according to the following enquiries:

- Technical Assets Management;
- First Level Maintenance;
- Planning and Safety;
- Data base;
- Maintenance work;
- Stocks;
- Cost Analysis;
- RAMS Analysis;
- RCM;
- RBM;
- Transversal Tools;
- 5S;
- TPM;
- LEAN Maintenance.

This approach corresponds to a diagnosis method of the maintenance state, centered in a questionnaires sequence, which answer evaluation identifies the positioning of the maintenance model of the organization, [4-5]. The methodology is based on the following phases:

- i. Data Collection;
- ii. Analysis of the collected data;
- iii. Establishment of an action plan of improvements.

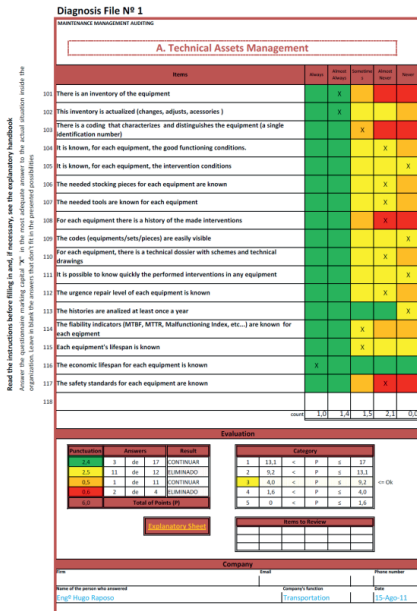


Figure 2.1 – Diagnosis File n°1



Figure 2.2 – Diagnosis File n°14

Perspectives	Items
Organization Basis	Management of Technical Assets; 1 st Level Maintenance; Planning and Safety; Data Base; Maintenance Work; Stocks; Cost Analysis.
Transversal Concepts	RAMS Analysis; RCM; RBM; Transversal Tools.
New Management Models	5S; TPM; LEAN Maintenance.

Table 2.1 – Questionnaires and items

2.2 THE QUESTIONNAIRES

The present methodology is based on fourteen questionnaires, called “Diagnostic Files”. There is a file for each item above referred, to be filled by the companies’ responsible technicians or by consultant technicians external to the company.

Items	Activity	Maximum Punctuation	Minimum Punctuation
1	Technical Assets Management	17	9.2
2	First Level Maintenance	16	7.2
3	Planning and Safety	18	8.6
4	Data base	12	5.0
5	Maintenance Work	16	6.9
6	Stocks	16	8.2
7	Cost Analysis	14	6.8
8	RAMS Analysis	18	9.5
9	RCM	14	6.4
10	RBM	14	7.0
11	Transversal Tools	12	5.4
12	5S	18	10.8
13	TPM	20	12.0
14	LEAN Maintenance	21	10.6

Table 2.2 – The 14 items of the model of diagnosis, with respective maximum and minimum punctuations

In case of doubt, each questionnaire is accompanied by an explanatory file regarding the presented questions. The Table 2.2 presents fourteen states that compose the diverse questionnaires in which is based on the model of diagnosis, with the respective maximum and minimum punctuations.

For each one of the referred items, it is elaborated a diagnostic file, enquiry type, with various questions and with five answer possibilities that are the following:

1. Always (always verified in the company);
2. Almost always (not always verified in the company);
3. Sometimes (sometimes verified in the company);
4. Almost never (few times verified in the company);
5. Never (never verified in the company).

2.3 EXPLANATORY FILES

With the goal of minimizing the doubts about the content and comprehension of the formulated questions in the diagnostic files, these are accompanied by an explanatory file, individualized by questionnaire that allows, question after question, to interpret them and knowing what answer option to mark.

1201	Separates the useful equipment, tools and materials from the useless, eliminating the unnecessary (Seiri)	
Explanation regarding the true		Explanation regarding the false
Separates the useful from the useless, eliminating the unnecessary and the work begins to be placed in order, so that it is only used what is really needed and applicable. For that, it is important to have the necessary, in the adequate and controlled amount to simplify the operations.		Does not separate the useful from the useless and does not eliminate the unnecessary.
Improvement Suggestion		
Separating the useful from the useless, eliminating the unnecessary and the work begins to be placed in order so that it is only used what is really needed and applicable. For that, it is important to have the necessary, in the adequate and controlled amount to ease the operations. It can also be interpreted as having the sense of Utilization, Arrangement, Organization, and Selection. It is essential to know how to separate and classify the useful objects and data from the useless through the following way: what is always used, place near the work place; what is almost always used, place near the work place; what is occasionally used, place a little away from the work place; what is seldom used but needed, place away, at a determined place; what is unnecessary must be reformed, sold or eliminated, as it occupies room necessary to the work objects. Advantages: reduces the necessity and spending on room, stock, storing, transport and insurance; eases the internal transport, the physical arrangement, the production control; avoids the purchase of components and materials in double and avoids also the damage to stored materials or products; increases productivity of the machines and people involved; brings more humanization, organization, economy sense and less physical fatigue and more ease to operate; decreases the accidental risks of the using of these materials by the personnel. All the team members must know how to differentiate the useful from the useless, what is really necessary from what is not.		

Table 2.3 – (L–5S) for each question, two explanations and an improvement suggestion

As it is obvious, it is necessary to have in account the company’s context. For example, if the company subcontracts are all new ones, it won’t be able to answer the questions about the realization of their works. In this case, should not be answered any question. If more than 25% of the presented questions on each file are not answered, then that file is not considered for the knowing of the final evaluation, because it is considered that there are not gathered the minimum conditions for a sustained evaluation of that item.

In table 2.3, it is presented the first question of diagnostic file “L”, regarding 5S. In addition to the questions it is also exemplified the given interpretation – both in the cases of the answer be true or false. At last, as it is illustrated in the table, for the exemplified question, the content of the answer for that question should be interpreted as an improvement suggestion.

The diagnostic files are identified in the upper part, through the item number of which they correspond to (1 to 14), by the designation “Diagnostic File no” and are identified by a heading that designates the corresponding item. It is followed by an intermediary zone with a grid with the questions and the columns reserved for each answer. Each one of the lines begins with the indication of a number associated to each affirmation, composed by three or four digits. The first one represents the number of the enquiry file and the two remaining identify the order of the affirmation. It is based on this numeration that can be searched help at the “Explanatory Dossier”.

In the following columns of the table it appears the five answer possibilities – Always, Almost Always, Sometimes, Almost Never and Never – according to what has been referred, should be answered one and only one option. In case of being impossible to respond, should not be answered any question.

At last, the bottom part is reserved to the determination:

- Of the obtained punctuation;
- Of the consequent classification by category;
- Of the achieved elimination criteria.

2.4 ANALYSIS OF DATA

According to the obtained answers of the questionnaires, that interpret the actual state of maintenance, each diagnostic file achieves determined punctuation that is divided into five categories: Category 1; Category 2; Category 3; Category 4 and Category 5. The category 1 means very good positioning of the company and category 2 a good positioning. The central or intermediate situation is between the positive and negative, and represents a reasonable positioning. Inside the negative situations, category 3 means that there are some aspects that need improvement in the organization, while category 4 means a bad accomplishment in terms of maintenance management, indicating that it would be needed a vast and deep reorganization.

So, for each item, category 1 is the category associated to higher punctuations, meaning good positioning of the company, while category 5 is related to the lowest punctuations, associated to a bad accomplishment.

Each one of the states must achieve a minimum punctuation in order to sustain the positioning of the following item, this is, the company cannot adequately guarantee the implementation of the questions of a certain item, without previous state have achieved a certain position that should be considered favorable. Practically, it is established that the company must achieve the threshold of the third category to guarantee the application of the next item.

2.5 ELIMINATION CRITERIA

To each answer possibility to the diverse questions, it’s attributed a degree of importance, working as an elimination criteria, according to four colors – green, yellow, orange and red – with the interpretation given in table 2.4.

Green	Adequate Answer This answer is always the desirable one.
Yellow	Inadequate Answer Only some answers should be this type, the company shall improve them.
Orange	Exceptional Answer Few answers should be this type, although these answers are not eliminatory, the company should improve them as soon as possible.
Red	Critical Answer The company should never have this type of answer, being the first to be reviewed.

Table 2.4 – Importance criteria of the answers in the positioning of the maintenance state

It is considered achieved an elimination criteria if the company answered some question of critical importance or if it overcame the maximum number of admissible answers related to questions of exceptional importance.

The evaluation of the questionnaires allows determining the state of the maintenance management in the company regarding the present management of their equipment and facilities.

It can be verified in diverse questionnaires that the columns related to the answers of the type “always” or “almost always” are the most desirable answer possibilities, so it is attributed to this type of option the “green” color.

In each questionnaire, for questions answered in a negative or central position, namely, “sometimes” or “almost never” or “never”, the model automatically elaborates a report of weak points (orange zone answers) or critical points reports (answers obtained in red zones).

2.6 ELIMINATORY GRID

The eliminatory grids options are red, orange, yellow and green colored cells, if they are critical, exceptional, inadequate and adequate, respectively, allowing identifying the maintenance management state of the company, according to the process previously described.

The color attribution process in the elimination criteria resulted of the importance of each question contributes to the maintenance state diagnosis and, consequently, its implications in the maintenance reorganization.

The punctuation (R) obtained by the company results from the following formula:

$$R = \sum R_S \oplus \sum R_{QS} \oplus \sum R_{AV} \oplus \sum R_{QN}$$

Being:

- R_S – “Always” Answer
- R_{QS} – “Almost Always” Answer
- R_{AV} – “Sometimes” Answer
- R_{QN} – “Almost Never” Answer

2.7 IMPROVEMENT PLAN

The last phase of the diagnosis allows establishing an improvement plan of the audited company. After the filling and evaluation of the punctuation of the different questionnaires, the model produces various positioning graphics, called “Radar Maps”, that allow the graphical viewing of the company positioning and the aspects of organizational improvement.

The diagnostic model automatically produces six radar maps of positioning:

1. General position of the company’s maintenance state;
2. Specific position of the company’s maintenance state;
3. Position of the company at the organization basis;
4. Position of the company regarding the transversal concepts and new maintenance management models;
5. Position of the company regarding the transversal concepts;
6. Position of the company regarding the new maintenance management models.

Beyond the positioning radar maps, could be elaborated diverse reports that allow making a characterization of the maintenance state of any organization such as:

- Positioning Report;
- Critical Points Report;
- Fragile Points Report;
- Weak Points Report.

The positioning report, as its name indicates, has the objective of positioning the maintenance and organization state at diverse

stages, considered in the diagnosis model, allowing them to make an immediate characterization of the general diagnosis of the maintenance along the radar map. Regarding the other reports, they allow to obtain the referred points according to their designations (critical, fragile and weak), as well as the improvement suggestions for each marked point.

3. MODEL APPLICATION

The Holistic Model for Maintenance Diagnosis State (HMMDS) was validated in three Portuguese transports companies. Besides these three companies, the HMMDS was also validated in a Portuguese company of the food sector that possesses an exportation rate of 20% of its producing. A transversal aspect in these companies is the fact that the maintenance activity is extremely important to the prosecution of their mission.

However, our focus was in the transportation sector, which study is shown in this point.

According to table 3.1, it can be perceived that the differences among different companies analyzed are significant, as it is expressed by the correspondent maintenance state diagnosis. It can be told that those differences are not because of their dimensions and structures but also of the maintenance’s politic practiced, what means that each one of them represents a different management model.

Item	Activity	Company I Punctuation	Company II Punctuation	Company III Punctuation	Company I Category	Company II Category	Company III Category
1	Technical Assets Management	9,6	16,1	17	3	1	1
2	First Level Maintenance	8,9	15,0	15	2	1	1
3	Planning and Safety	8,6	15,6	16,5	3	1	1
4	Data base	7,1	11,0	12	2	1	1
5	Maintenance Work	6,2	15,2	14,5	3	1	1
6	Stock	12,1	13,6	14,50	2	1	1
7	Cost Analysis	6,8	10,7	13,10	3	1	1
8	RAMS Analysis	7,6	15,5	9,5	3	1	3
9	RCM	2,2	3,6	5,6	4	3	3
10	RBM	2,0	6,3	6,8	4	3	3
11	Transversal Tools	0,3	2,6	4,8	4	3	3
12	SS	4,2	8,7	10,6	4	3	3
13	FPM	3,1	4,3	10,7	4	4	3
14	LEAN Maintenance	1,4	4,5	10,3	5	4	3

Table 3.1 – Comparative table of the achieved punctuations – Companies I, II and III

They are obvious the big differences among company I and companies II and III, because they are in an upper level, regarding their maintenance state, as they represent organization levels similar to international companies considered as top references. Regarding the differences between company II and company III, it is noted that the diverging points are related to the application of on condition maintenance, because the company II applies it in a more systematic way. The company III has a better punctuation at most of the items.

Concluding, it can be said that companies I, II and III should improve their maintenance state, at different levels, for what it is suggested that they may introduce new techniques and methodologies, in order to improve their maintenance management.

The figure 3.1 illustrates the global radar maps of each of the diagnosed companies.



Figure 3.1 - Radar Maps

4. CONCLUSIONS

The diagnosis model, discussed in this paper, was named as Holistic Model for Maintenance Diagnosis State (HMMDS), because it allows the analysis and the understanding of the situation of the maintenance management state in the organizations. The HMMDS permits emphasizing the strong and the weak points of maintenance that is made and allowing the identification of the weak points to improve in the future.

The method is based on fourteen questionnaires, grouped into three perspectives: organization basis; transversal politics; and new management models.

The HMMDS was validated in three transports companies and one company of the food branch, having been showed its suitability and versatility.

The application of HMMDS in the transports sector allowed concluding that the analyzed companies need to develop the maintenance management that is carried on, because only that could potentiate its production.

The HMMDS is a support tool to the maintenance management that is available to all organizations, being generalist, adaptable and evolutionary according to the new challenges of maintenance or, more generically, of the new challenges that the technical assets management place in action.

5. REFERENCES

- [1] Farinha, J. M. T. (2011): *Manutenção – A Terologia e as Novas Ferramentas de Gestão*. Lisboa: 1ª Edição, Monitor – Projecto e Edições, Lda.
- [2] Oliveira, D. P. R. (2006): *Planejamento Estratégico: conceitos, metodologia e práticas*. São Paulo: Atlas.
- [3] Certo, S. e Peter, P. J. (1998): *Administração Estratégica*. São Paulo: Makron Books
- [4] Costa, J. A. M. (2002): *Uma Abordagem ao Diagnóstico do Estado da Manutenção em Empresas Industriais*. FEUP, Porto, Tese de Mestrado.
- [5] Costa, J. A. M.; Farinha, J. T.; Vasconcelos, B. C. (2000): *Diagnóstico da Manutenção – Um Passo Fundamental na Sua Reorganização*. Revista MANUTENÇÃO, Nº 65, 2º Trim. 2000, pp 14-18.

6. COMPLEMENTARY REFERENCES

- Assis, R. (1997): *Manutenção Centrada na Fiabilidade*. Lisboa: Lidel – Edições Técnicas Lda.

- Cabral, J. S. (2006): *Organização e Gestão da Manutenção*. Lisboa: 6ª Edição, Lidel – Edições Técnicas Lda.
- Charrier, J.C.; Kemoun, K. (1990): *Organização e Modernização Industrial*. Lisboa: Lidel – Edições Técnicas Lda.
- Dias, J. M. R., (2003): *A Gestão da Manutenção em Portugal*. UTL e ISEG, Lisboa, Tese de Mestrado.
- Ferreira, L. A. (1998): *Uma Introdução à Manutenção*. Porto, Publindústria.
- George, M.; Rowlands, D.; Kastle, B. (2004): *What is Lean Six Sigma*. George Group: MacGraw-Hill.
- Leocádio, M. N. S. (2004): *Sobre a incorporação RAMS no desenvolvimento de Produtos de Base Tecnológica: Uma abordagem Holística a Veículos Ferroviários*. Dissertação de Mestrado em Engenharia a Gestão de Tecnologia. Instituto Superior Técnico, Lisboa; Portugal.
- Levitt, J. (2008): *Lean Maintenance*. Publisher: Industrial Press, Inc. ISBN-10: 083113352X. ISBN-13: 978-0831133528.
- Loureiro, H.(2007): *Excel 2007 Macros e VBA*. Lisboa: 3ª Edição, Lidel – Edições Técnicas Lda.
- Marimba, A.; Farinha, T. (2003): "Controlo e Monitorização da Gestão da Manutenção: Metodologia de Auditoria Interna Permanente". Resumo da comunicação nas actas das 3ªs Jornadas Politécnicas de Engenharia. Pp65. ISBN 972-99064-0-8. Comunicação em CD, ISBN: 972-99064-1-6.
- Monchy, F. (1989): *A Função Manutenção*. França: Ebras/Durban – Editora Durban Lda, Editora Brasileira Lda.
- Nakajima, S. (1986): *Introdução ao TPM*. São Paulo. IMC Internacional Sistemas Educativos. Tradução: Nishimura, Mário (1989).
- Ohno T. (1988): *Toyota Production System: Beyond Large – Scale Production*. New York, ISBN: 0-915299-14-3.
- Ohno, T. (1997): *O sistema Toyota de produção: além da produção em larga escala*, tradução de Cristina Schumacher. Porto Alegre: Artes Médicas.
- Osada, T. (1991): *The 5S's: Five Keys to a Total Quality Environment*. Quality Resources. ISBN-10: 9283311167. ISBN-13: 978-9283311164.
- Pereira, F. S.; Lopes, C. (1996): *Aplicação de um Projecto TPM na Fábrica de Papel*. Actas do Quinto Congresso Nacional de Manutenção Industrial - APMI, Figueira da Foz.
- Pinto, C. V. (2002): *Organização e Gestão da Manutenção*. Lisboa: 2ª Edição, Monitor – Projecto e Edições, Lda.
- Shingo S., (1985): *A Revolution in Manufacturing, the SMED System*. Productivity Press, Cambridge, Massachusetts and Norwalk, Connecticut.
- Silva, R. P. (2004): *Gerenciamento do Setor de Manutenção*. Monografia (Pós Graduação, Gestão Industrial) – Pró Reitoria de Pesquisa e Pós-Graduação, Universidade de Taubaté, Taubaté.
- Simões, J. A. M. (2011): *Indicadores de Performance em Manutenção Industrial Utilização, Valor e Disponibilidade de Informação*. UC, Coimbra, Tese de Mestrado.
- Souris, J.P. (1990): *Manutenção Industrial – Custo ou Benefício*. França: Lidel – Edições Técnicas Lda.
- Souza, R. G. F. (1999): *Desenvolvimento do Sistema de Implementação e Gestão da Manutenção*. Porto alegre, Pós-Graduação na U.F.R.G.S., Brasil.
- Wireman, T. (1991): *Total Productive Maintenance*. New York: 1ª Edição, Industrial Press Inc.

Lean versus Kanban – An Actual Paradigm?

Pedro Dias
CEMUC
Coimbra
Portugal

+351 239 790700

jesus.pjd@gmail.com

José Torres Farinha
CEMUC
Coimbra
Portugal

+351 239 790700

torres.farinha@dem.uc.pt

Inácio Fonseca
CEMUC
Coimbra
Portugal

+351 239 790700

Inacio.fonseca@dem.uc.pt

Luís Andrade Ferreira
FEUP
Porto
Portugal

+351 22 508 14 00

lferreir@fe.up.pt

ABSTRACT

In a manufacturing industry it seems to have an eternal conflict that is the lowest possible levels of stocks, the maximum levels of availability of equipment and the total elimination of wasteful that is an equation of difficult solution.

Nowadays there are a lot of methodologies, technologies and algorithms to reach this objective but, in fact, through all Europe we continue observing very important industries dealing with the above dilemma.

The paper presents the solutions that are being implemented in a manufacturing industry of a multinational company that tries to optimize the Kanban system in the production process, the JIT and the Lean, describing the problems and solutions found.

The present industry produces high quality products and, through benchmark and the deep knowledge of their processes, it knows that it is possible to optimize a lot their efficiency similar to its principal competitors. However, it needs to make a profound changing in human resources, some technological improvement and, at last, to give a step ahead in direction into the best competitors in the world.

The industry believes that the step necessary to reach that objective is implementing the trinomial Kanban-JIT-Lean but is dealing with some internal cultural problems. These problems may be similar to many others industries and it is because of this that this paper can be understood as a case study that could be used by any other companies.

Keywords

Maintenance, 5S, Lean, Kanban.

1. INTRODUCTION

Nowadays industrial reality, it is crucial that companies are up-to-date with the most adequate management methodologies. The continuation of the use of old management concepts makes companies obsolete and doomed to failure, especially where competition is high. For a company to succeed in the market, to be competitive, it needs to produce quality products and services at lower costs possible. With competition from emerging markets (the case of India and China, for example), where production costs are much lower than in Europe, it is necessary that the quality be the engine of the company's sales.

Producing well, deliver products on time, and answer within a time attractive to the customer, are crucial aspects to compete with the new world competitors.

The new management methodologies, helped Japanese companies to reach the top of success, displacing old technologies and

ideologies, both American (as is the case of Toyota versus General Motors), Swedish (as is the case of Samsung versus Nokia), or Germanic; so, it is necessary to follow new technological developments and to be always on top on of the international management methodologies.

The industrial production and the maintenance departments, in many cases, have difficulty to work side by side. Blaming each other, try to justify their failures. The production department blames the maintenance department because it does not perform the actions necessary for the proper functioning equipment; on the other hand, the maintenance department blames the production department by a bad tuning of equipment, a cleaning and a defective negligent operation by operators.

In this "game", the maintenance costs increase, what requires more hand labor for the maintenance of equipment and at the end, the whole team spends his time "extinguishing fires". In addition, production costs increase, because the downtime of equipment leads to increased overtime to meet delivery deadlines. This whole storyline may cause a financial collapse in an industrial unit, leading to a lack of competitiveness and a future bankruptcy.

In view of the elimination of these problems, they were developed several management methodologies, both to be applied to maintenance activities and production activity, to provide a system "thin", in the costs of maintenance and production, requiring coordination between these two departments with the objective to eliminate dead times and repair times.

From the large amount of methodologies disseminated throughout the world, this paper discusses three approaches that are considered to be fundamental for a company's success, that are the followings:

- JIT (Just In Time);
- Kanban;
- Lean.

2. STATE OF THE ART

The new management methodologies, like JIT, Kanban, and Lean, are well known both the academic and professional community. By consequence, there are a lot of bibliography exploring theoretically and case study applications of these new methodologies. Explains the most part of these methodologies and tools in order to maximize the organizations efficiency [1].

Just In Time, is a management methodology originated in Japan in the 60's, in Toyota company' that had as great mentor Mr. Taiichi Ohno. This methodology is a management philosophy that seeks

to eliminate any kind of waste, which is based on three basic ideas:

- Integration and optimization of the process as a whole, eliminating everything that does not add value to the product;
- The application of Kaizen system, ie, the search for continuous improvement, encouraging the development of human potential;
- Understanding and respond to customer needs, eliminate preliminary, intermediate and final stocks, and trying to produce in smaller times and right at the first time, in order to eliminate production costs in interim and final inspections.

In Figure 1, it is possible to summarize the essence of JIT, where the PULL system is crucial for a smooth production with low stocks, [2].

The Kanban method is one more methodology that came from Japan. The translation of the word is, literally, "Registration or Visible Plate."

This methodology is commonly used to automate the production orders. It is placed a proper amount of letters where they are flipped or drawn, will give order to the production machine or production cell. In this system, which order production is always who uses it, ie, the final needs will enable the successive letters for upstream production.



Figure 1. Just In Time

This methodology has three variants: E-Kanban, Kanban Production, and Motion Kanban.

1. E-Kanban - also called Electronic Kanban - is the application of traditional kanban system, but replacing the order letters for electronic indicators. The calculation system works of the same way, but has the benefit do not have the risk of loss of letters during the filling process or order to fill the container. It can be seen in Figure 2, where an electronic signal does start a new production, having always a car passing to bring the existing production and giving order for new production, [3].
2. Kanban Production - is the best known and most common. It is used a card set with a specific amount of

product. Always this amount is consumed, this card returns to the manufacture there by giving a manufacturing order. Figure 3 shows how a Kanban frame can work, where a system of letters when are changed to an order frame, can cause restarting of production of a certain reference, [4].

3. Motion Kanban - also called transport Kanban - in this case, we use the called Kanban car. This methodology consists in the existence of a car with several Kanban letters, which give order to the manufacturing cell that fills the car according to the number of cards available. After the production of the car, it will move to the next process that will consume all its contents. This passage includes the return of another car to manufacture, giving order for new production. In Figure 4, it is represented the kanban car system, where there is only one new production order when there is a car back empty, [5].

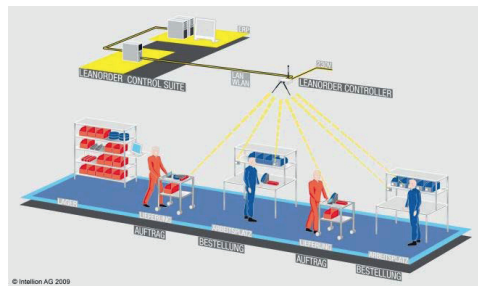


Figure 2. E-Kanban system

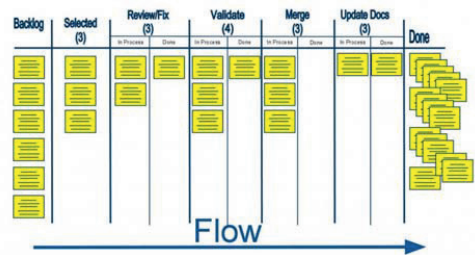


Figure 3. Kanban production system

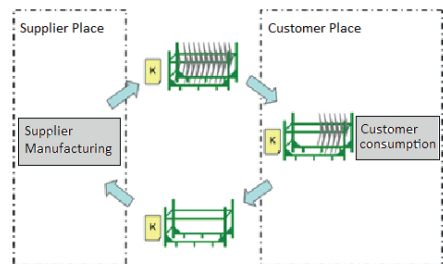


Figure 4. Motion Kanban system

About Kanban systems, its design and implementation see also [6-7].

Lean, also called Lean Thinking that, translated literally means "Lean Thinking". Like other methodologies, Lean also born in Japan, more precisely in the company Toyota Motors Corporation. It was based on other methodologies already developed by Toyota such as the TPS (Toyota Production System), TPM (Total Productive Maintenance) and JIT (Just In Time), among others. This methodology allowed that Toyota take off the place of General Motors from the top of the automobile industry.

As the word means, this approach tends to eliminate everything that is waste, that is, to eliminate any process that does not add value to the product. When setting value it means to classify any equipment, motion, stock, etc. that it is unnecessary to produce the product. These points only increase the production costs and are not adding anything new to the final product consumed by the customer.

In Lean philosophy, all these processes should be eliminated, thereby increasing production at lower cost.

These wastes have been identified in TPS methodology and are the following:

1. Production excess;
2. Waits;
3. Transport and movements;
4. Waste from the process itself;
5. Stocks;
6. Defects;
7. Unnecessary works.

Analyzing these points, a company to be competitive has to produce well at first, thus eliminating inspections, rework, etc..

Looking at what it is necessary to change, it seems easy, but, in fact, this methodology encompasses much more than that. For a practical and efficient implementation, it is necessary that the entire business community (from the administrator to the operator) is in tuning and have an open mind to accept the changing. Only with this general involvement of all is possible switching techniques and ways of working long rooted in the company.

Some authors define seven principles for applying Lean in a business unit that are shown in Figure 5.

In Figure 5, we can conclude that any company should: to know their customers and identify what they want (as well as all stakeholders); to identify the needs for the production of the product, eliminating all that does not add value (value concept); to define clearly the previous item, identifying all the value that each post adds to the product (concept of value chain); to identify a production flow, in a simple and fast connection of the process, from start to output the final product (flow optimization); to determine needs among all the stations of production, with the objective of do not create intermediate stocks, and achieving that the needs are always identified by stations downstream making that production is not endangered for lack material (pull system); to seek an harmonization in the plant, in order that, from the operator to management team there is a conjunction to make the best looking (perfection); finally, always look for improvement,

identifying the fails, as well as the ways to improve (always innovating), [8].

About Lean methodologies see also [9-10].

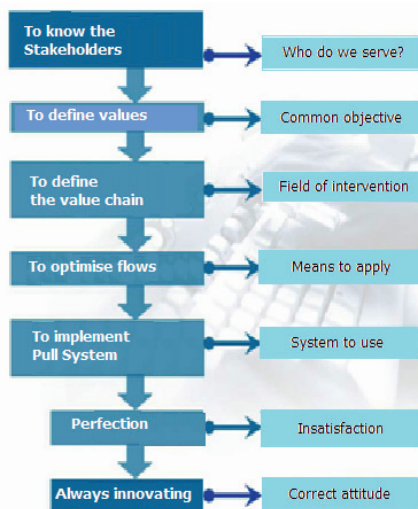


Figure 5. Seven principles of Lean Thinking

3. METHODOLOGIES APPLICATION

The methodologies above mentioned were born with the intention of optimizing the flow of a productive plant. These methodologies have reached very good results, achieving, each one by itself, better flows and generate added value at lower costs.

For an application of the system, it is necessary, first of all, that all flows be optimized in order to have a standardization of the entire production system. For this standardization it is necessary to apply two systems: Kanban system; and leveling system of production, also called Heijunka.

3.1 Industrial Methodologies Application

3.1.1 JIT Industrial Application

The need for production leveling results from the importance of eliminating the intermediate stocks. Through the application of a Pull system, ie, a process pulled from the client to the first level of manufacture, in which all flows of intermediate production must be working with well-defined times in all phases, in order do not cause shrinkage of raw materials and intermediate products.

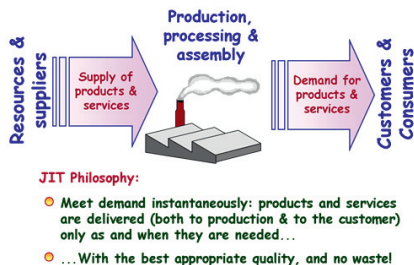


Figure 6. Flow Just In Time

As shown in Figure 6, from the supplier to the customer's delivery, the entire system is under control. For the reduction, or even elimination, of intermediate stocks, it is necessary to perform the calculation of the takt-time, i.e. the cycle time necessary to create a constant flow determined by Pull. This process time is calculated using the following formula, [11]:

$$Talk\ Time = \frac{Available\ Time}{Search\ on\ the\ Available\ Time}$$

This point requires that the manufacturing cycle time does not exceed the Takt-time, in order do not to cause delays in deliveries, so there is no waste in the process.

In summary, for the implementation of JIT in this context, it is essential that the process is controlled and stabilized, and that there is monitoring of the entire product, from raw material input to tie off the finished product, and that this process is triggered by the customer, eliminating intermediate stocks and the finished product.

In the maintenance activity, these concepts are also applicable, especially in the management of material resources. Being the client the maintenance sector, the management can be performed by Pull system, ie, through the application of planned maintenance, predictive and on-condition, it is possible to determine the evolution of equipment condition variables, in order to minimize downtimes and meet the production leveling analyzed before.

3.1.2 Kanban Industrial Application

The Kanban system can be seen as the application of the pull, in which all the work orders are indexed from the customers' requests for the manufacturer. In case of a manufacturing cellular chain, all production orders or assembly line, are given automatically by the outflow from end cell to the initial manufacture cell.

As discussed in the previous paragraphs, the application of JIT system aid Kanban potentiate the autonomy as the order to manufacture is launched.

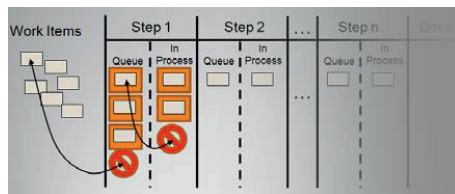


Figure 7. Kanban board with queue and process cards

3.1.3 Lean Industrial Application

As mentioned initially about Lean system, this implies a comprehensive methodology that includes methods of rational management and, in particular, the called "thin".

To achieve the implementation of this concept, it is necessary that the company has already some methodologies implemented, such as, for example, 5S. The 5S correspond to a method for organizing and cleaning, with the following meaning, [11]:

- i. Seiri (Sense of Organization) – to separate the useful from the useless; to identify unnecessary things in the workplace.
- ii. Seiton (Sense Storage) - set a place for everything; check that everything is in its right place.
- iii. Seiso (Clean Sense) - split the job and assign a zone to each member of the group; to clean every area of the workplace, as well as the surrounding areas.
- iv. Seiketsu (Sense of Standardization) - set a standard of general housekeeping for the job; identify visual aids and procedures, and standards of housekeeping and cleanliness that result well; normalize throughout the factory, the equipment / job places of the same type.
- v. Shitsuke (Sense of Self-Discipline) - practicing the principles of organization, systematization and cleaning; eliminate variability, ie, always do well at first.

Figure 8, illustrates the above concepts, where all functions require a self-discipline to be adopted every day, [12]. For a Lean application it is necessary that all those processes be implemented, because they are the starting point for improvement.

Another key point is the generalization of its implementation, that is, there must be cooperation between production (operators) and the top management (leadership), ie, there must be communication among the community, in order that all changes be approved by everyone. Having the factory clean and organized, the next step is to identify flows and correct deviations, if they exist, as a way of improving. At the end of one bout (eight hours of work), there are many times that are wasted in movements that are considered expendable; to solve this problem, it is important to make a process optimization.

Another essential point is the determination of all material and equipment that are not necessary or which are inoperable and remove them. The result of this process is a company organized and cleaner, where what exists on the factory is only the necessary for its function. In the maintenance activity the methodology under discussion can be widely applied. Some of the concepts follow the same process as in production activity.

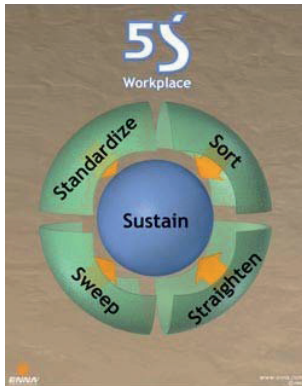


Figure 8. 5S plant

For example, by implementing maintenance plans and controlling the interventions, it can be determined the maintenance times in order to act with an adequate preparation. Conducting a survey of all the necessary tools for each piece of equipment and parts and normalizing measures, it is possible to save costs in these resources.

Subsequently, with the advance in interventions planning, it is possible to determine the spare parts and tools needed and, by consequence, to eliminate paths that would be required to fetch some tools or spare parts that were not taken initially for the location of works.

4. LEAN-KANBAN IMPLEMENTATION IN AN AUTOMOBILE INDUSTRY

The management models above exposed implies the termed "thin" management, because they seek to eliminate stocks and minimize production costs. In terms of reducing costs, especially in the aspect of production, these concepts are particularly applicable in the automobile industry. With an optimization process, an improvement in the exchange of tools and references can increase a lot the company's competitiveness.

One of the weak aspects of the automobile industry is the disposal of stocks. In this industry, at the level of the assembly lines, has a very high oscillation in the production and to achieve answers to this change request, there must be a good reservation buffer of finished material, otherwise it may result in lack of productive capacity of the company.

Another significant problem which affects the production indices of the company is designated sub-assembly that corresponds to the pre-assembly processes that are essential for placing products on assembly lines to make the final product. Being these devices often the starting point for several references, it is necessary to create a buffer which is able to feed all cells uniformly and not causing any stops for lack of material.

In the automobile industry it is common to use the Kanban system card, from which is calculated one base stock which, when activated a request for production, the equipment has the ability to produce the reference before the cell need it again.

In this system, there is a counterbalance system for the JIT, because there is the need of having intermediate stock, which is something in system JIT to eliminate.

To calculate the intermediate stocks they are taken into consideration various concepts such as: the number of cards kanban of reserve, which must correspond to the delay time necessary for make the changes in the machine to produce the new reference; to have a stock of reserve corresponding at a certain time of use of the cell, i.e., the order of repair is activated when, for example, the stock quantity corresponding to a previously calculated corresponding to one day's working cell. By this way, it is ensured that the team corresponding to that production unit has 24 hours to produce the corresponding amount to Kanban card that exist as buffer.

4.1 Balance Lean-Kanban in an automobile industry

In the automobile sector, with the present huge sales oscillation, it becomes very difficult to accurately calculate the amount of Kanban cards or even the necessary buffer. By default, the company taken as a reference in this paper uses a system of five days' finished stock product, thus responding adequately to customer requests.

But the finished product is not the unique problem, but along the production chain. The intermediate stocks are high compared with the homologous business and the need to optimize the process with regard to internal amounts of sub-assembly is extremely urgent.

The company used as reference in this paper is in a process of implementing Lean methodology in its production process. The company began by identify the problems and finding the best solutions for them. One of the first major issues to be resolved was the workflows, given that there were many crossings and certain operators needed to walk large distances to transport sub-assembly products. To address this issue, we proceeded to a reformation of the internal circuit production, modifying it to a more linear process, from raw materials to feed the first cell and then in a simple way runs directly throughout the production circuit to the last production cell. In this circuit it was necessary to install dynamic shelves in order to be able to feed the product from an entry point and, by friction, this goes sliding until the output side, without the operator ever having to violate the space of other coworker. In this change, we have eliminated many unnecessary movements which contribute only to decrease the cycle time of production and the cost of the final product.



Figure 9. Primary Layout of the Production Process

Figure 9 shows the initial flow prior to the Lean optimization. It can be seen that the flow is not constant, with some advances and retreats to the final product.

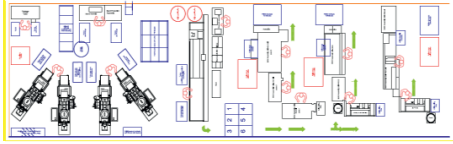


Figure 10. Final Solution of the Production Process

In figure 10 it is shown the changes made in the production flow.

Another very important point of improvement was the optimization and control of the production process. Taking the extreme the JIT methodology, we started by eliminating anomalies contained in the production process to achieve its stabilization, producing well from the beginning. This last aspect is still not fully resolved, but it is a crucial stage for their final consolidation.

In figure 11, it can be seen the practical application of what is shown in Figure 10 and as a Lean methodology can contribute to excellence in respect of productive organization. Other points that are under discussion to be applied in the near future relate to the normalization and optimization, as regards to the exchange of tools. Making this task as quickly as possible and streamline as much as possible all exchanges of reference in order to monetize more and more the equipment and optimize the SMED (Single Minute Exchange Die). At this point, the combining of these features with Kanban card system, permits determining the optimum amounts to produce.



Figure 11. Last Flow Production

With these aspects to be followed and applied, you can move to a higher level of management, and implement an appropriate interface with the maintenance activity. This interaction allows the program to conjugate the planned interventions with production, to monitor the evolution of condition variables, including wear parts and, finally, to optimize the maintenance sector.

5. CONCLUSIONS

This paper showed how it was possible to improve the flow at the level of the production process in an industrial plant, through the introduction of methodologies such as 5S, Kanban and Lean. This improvement resulted in an increased customer satisfaction

because of their approval about the changes made; it was possible to see how easily it was perceived the progress made in the production process. All the steps of the product manufacturing became easy to identify. The involvement of all cells operators became effective, including their participation through their suggestions and proposals for improvement. While they are working in a motivating environment all the time working become more productive.

It was possible to increase production by 80% (initially the total production per shift was around 3.300 pieces; nowadays it is around 6,000 pieces per bout), minimizing some dead times with the application of new equipment and eliminating about 40% in the distances between cells.

This increase of the production was possible with implementation a new production cell. The construction of this cell was an important contribution, because it make possible to increment a lot the cadence of production. However, the improvement on the actual system of production and the new equipment, made possible to reach this value.

But, to reach this value it is necessary that there exists product in the previous cell. The change from previous production system to present flow based on kanban system made possible to control adequately the production and prepare quicker the change of references.

With all these changes it was possible to identify in an easy way the correct implementation of all methodologies, especially the 5S.

Another aspect that aroused some interest was the control of the amount achieved of saving scrap. With these methods applied and the requirement to do well at first (giving the operator a comfortable working environment and pleasant) it was able to eliminate an intermediate inspection post and decrease by about 70% the scrap produced in the startup process.

With the application of racking it was still possible to delete a constant planning, and given the production order through the system pull.

REFERENCES

- [1] José Manuel Torres Farinha (2011): *Manutenção – A Terologia e as Novas Ferramentas de Gestão*. MONITOR, Lisboa, Portugal. ISBN 978-972-9413-82-7.
- [2] <http://www.qualitycoach.net/poster/jitennaposter.htm>. (acedido em 15 de Setembro de 2012).
- [3] <http://www.infoq.com/articles/agile-kanban-boards>. (acedido em 28 de 09 de 2012).
- [4] <http://www.qualitycoach.net/poster/jitennaposter.htm>. (acedido em 15 de Setembro de 2012).
- [5] <http://pt.encydia.com/es/Kanban>. (acedido em 20 de 09 de 2012).
- [6] Akturk, M. S.; Erhun, F. (1999): An overview of design and operational issues of kanban systems *International Journal of Production Research*. Volume 37, Issue 17.
- [7] Sugimori, Y.; Kusunoki, K.; Cho, F.; Uchikawa, S. (2007): *Toyota production system and Kanban system - Materialization of just-in-time and respect-for-human*

- system. *International Journal of Production Research*, 15:6, 553-564.
- [8] <http://www.intellion.com/Architektur.409.0.html>. 2011. (acedido em 23 de 09 de 2012).
- [9] Lamming, Richard (1996): Squaring lean supply with supply chain management. *International Journal of Operations & Production Management*. Vol. 16, Issue 2, pp.183 – 196.
- [10] Cochran, David S.; Eversheim, Walter; Kubin, Gerd; Sesterhenn, Marc L (2010): The application of axiomatic design and lean management principles in the scope of production system segmentation. *International Journal of Production Research*. Volume 38, Issue 6. Pp1377-1396
- [11] Elsevier.<http://www.warehouseinc.ph/?p=8>. 2009. (acedido em 20 de 09 de 2012).
- [12] Pinto, João Paulo. *Pensamento Lean*. Lidel, Edições Técnicas, 2009.
- [13] <http://blogs.mulesoft.org/implementing-kanban-for-sustaining-engineering/>. 2009. (acedido em 24 de 09 de 2012).
- [14] <http://www.qualitycoach.net/poster/5sennaposter.htm>. 2010. (acedido em 28 de 09 de 2012).
- [15] Plenert, Gerhard Johannes (2007): *Reinventing Lean: Introducing Lean Management Into the Supply Chain*. Butterworth-Heinemann.

Author index

Abdul Razak, Jamin	101
Ahmadzadeh, Farzaneh	125, 131
Al-Douri, Yamur K.	197
Aliwan, Mustafa	95
Aljumaili, Mustafa	171, 191, 197
Andersson, Patric	37
Ang, Hermanto	89
Armas, Fátima	49
Asim, Taimoor	113, 119
Asplund, Matthias	15
Ball, Andrew	95
Bang-Steinsvik, Tor Ole	89
Barbosa, Fernando Maciel	219
Bergquist, Bjarne	145
Börcsök, Josef	175
Campos, Jaime	57
Candell, Olov	33, 43
Dias, Pedro	235
El-Thalji, Idriss	137
Famurewa, Stephen M.	15, 211
Farahani, Rezsa	69
Farinha, José Torres	49, 219, 229, 235
Ferreira, Luís Andrade	49, 229, 235
Fjellner, Jonas	211
Fonseca, Inácio	49, 219, 229, 235
Fransson, Torbjörn	33
Funk, Peter	63
Galar, Diego	23, 211
Gopinath, Rajesh	181
Gopinath, Subramani	101
Gu, Fengshou	95
Heyns, Stephan	205
Heyns, Theo	205
Jantunen, Erkki	137

Kabitzsch, Klaus	85
Karim, Ramin	23, 77, 167, 171
Kent, Stephen	1
Kour, Ravdeep	167
Krause, Jakob	85
Krini, Ossmane	175
Kumar, Uday	107, 181
Larsson-Kråik, Per-Olof	161
Lemma, Yonas	211
Lindsund, Isabelle	161
Lipsett, Michael G.	69
Mahmood, Yasser	171
Markeset, Tore	89
Mirangwanda, Gilbert	119
Mirzazadeh, Abolfazl	125
Mishra, Rakesh	113, 119
Morant, Amparo	23
Nadakatti, Mahantesh	107
Ngwangwa, Harry	205
Norrbin, Per	153
Oliveira, Rúben	49
Olsson, Tomas	63
Orand, Seyed Mostafa	125
Palo, Mikael	161
Parida, Aditya	7, 107
Park, Kyoo-Seon	113, 119
Pradhan, Suman	113
Rantatalo, Matti	15
Raposo, Hugo	229
Roberts, Clive	1
Rusu, Marius	1
Safari, Jalal	131
Schunnesson, Håkan	211
Shahzad, Atif	113, 119
Simões, António	219
Singh, Sarbjeet	107

Stenström, Christer	7
Söderholm, Peter	145, 153
Thaduri, Adithya	181
Tinney, Timothy	43
Tretten, Phillip	23, 77
van der Westhuizen, Stephanus	205
Wandt, Karina	7, 77, 191
Verma, Ajit K.	181
Vinod, Gopika	181
Wijk, Olle	37
Xiong, Ning	63

Organised by:
Division of Operation and Maintenance Engineering,
Process IT Innovations

Organizers:
Prof. Uday Kumar, General chair
Dr. Ramin Karim, Scientific chair
Dr. Aditya Parida, International chair
Mr. Anders OE Johansson, Local chair
Dr. Alireza Ahmadi, Programme chair
Dr. Philip Tretten
Mr. Anders Jonsson

www.emaintenance2012.org

ISBN 978-91-7439-539-6

