



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

Proceedings of

The 3rd international workshop
and congress on eMaintenance

June 17-18
Luleå, Sweden

eMaintenance

Trends in technologies & methodologies,
challenges, possibilities and applications



LULEÅ
UNIVERSITY
OF TECHNOLOGY





eMaintenance 2014

The 3rd International Workshop & Congress on eMaintenance

Luleå University of Technology, Sweden
17-18 June 2014

Supported by:



*Operation and Maintenance Engineering
Luleå University of Technology*

ISBN 978-91-7439-972-1 (print)
ISBN 978-91-7439-973-8 (pdf)

Luleå 2014
www.ltu.se

Available at: <http://pure.ltu.se/portal/en/>

Editorial

eMaintenance with application of information and communication technology (ICT) and condition monitoring is becoming the central theme for industry world over. The emergence of eMaintenance enabled with software, hardware, embedded sensors and wireless network are going to play a major role in the life cycle of the plant and machineries. eMaintenance concept is finding its applications in various industrial sectors like; railway, energy, mining and process industries. The focus of industry's eMaintenance applications is to monitor, control and carry out real time decision making from any location, with the goal to be dynamically competitive global players. The role of condition monitoring, and ICT in eMaintenance research and application is well established, which facilitates effective management of the operation and maintenance process. Thus, more and more eMaintenance applications with R&D focus for finding innovative solutions for the industries are at its prime today. eMaintenance facilitates and supports different interconnected technical and organizational levels in organizations by providing an effective and efficient infrastructure for decision-making.

Under this business and global scenario, the Division of Operation and Maintenance Engineering, who are involved in eMaintenance research projects since 2004, took a lead role to provide a platform for sharing knowledge and experiences amongst the academia and industry participants. This resulted in organizing the First International Workshop and Congress on eMaintenance in Luleå during June 2010. The Second International Workshop and Congress on eMaintenance was organized during 2012. After successful conduct of first two international workshop and congress on eMaintenance, the Third International Workshop and Congress is being organized during 17-18 June 2014, in Luleå University of Technology with a theme of "Trends in technologies & methodologies, challenges, possibilities and applications.

We received a good support and response from our industry partners, besides the academia in terms of technical papers and participation. The selected papers are compiled in this proceeding of the congress. The editors would like to thank all authors, who have contributed with their articles to this proceeding, and all reviewers for their support in reviewing the papers in time.

We hope, the participants and readers will find this proceeding interesting and useful in pursuing eMaintenance applications in the future!

Editors



Uday Kumar



Ramin Karim



Aditya Parida



Phillip Tretten

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 lung
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 Kråik
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 Ravdeep Kour
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



INFORMATION MANAGEMENT BY SAAB

For 75 years we have gained knowledge of how to develop, maintain and share product and logistics data.

This includes data for systems whose lifecycle spans over more than 30 years. Over the years the optimal use of information has become one of the main success factors for cost efficient design, manufacturing and operation.

With modern information & communication technology we today have the knowledge and tools to create efficient information solutions with a true life cycle perspective.

WHEN IS INFORMATION MANAGEMENT OF ESSENCE?

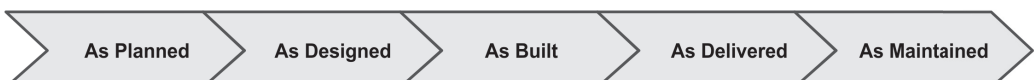
- In development, deployment and use of complex technical systems
- Stringent requirements on dependability, safety, and cost
- Long life-cycles and frequent updates and modifications
- For collaborative design and development

THE SAAB WAY

- A true life-cycle perspective
- 75 years of design and operation of complex systems of systems.

WE OFFER

Efficient Information Logistics Services for advanced Operation and Maintenance Solutions



The hands-on experience

eMaintenance LAB



Contents

eMaintenance 1A	1
• Top-of-Rail Friction Measurement of the Swedish Iron Ore Railway Line.	3
Yonas Lemma, Matthias Asplund and Matti Rantatalo	
• Analysis of Gauge Widening Phenomenon on Heavy Haul Line Using Measurement Data	9
Stephen M. Famurewa, Matthias Asplund and Paul Abrahamsson	
• Context Awareness and Railway Maintenance	17
Roberto Villarejo, Diego Galar, Carl-Anders Johansson, Manuel Menendez and Numan Perale	
eMaintenance 1B	25
• A Coloured Petri Net Model of Fleet Cannibalization	27
Jingyu Sheng and Darren Prescott	
• Attribute Control Chart Development by Evidential Reasoning	37
Fattaneh Javadi, Farzaneh Ahmadzadeh and Abolfazl Mirzazadeh	
• Detection and Diagnosis of Broken Rotor Bar Based on the Analysis of Signals from a Variable Speed Drive	41
Djoni Ashari, Mark Lane, Fengshou Gu and Andrew David Ball	
eMaintenance 2A	47
• Failure Prediction of Tidal Turbines Gearboxes	49
Faris Elasha, David Mba and Joao Amaral Teixeira	
• A Web Service-based Toolbox for Machine Diagnostics Based on Statistical Analysis	55
Luca Fumagalli, Simone Pala and Marco Macchi	
• Investigating the Effect of Water Contamination on Gearbox Lubrication based upon Motor Control Data from a Sensorless Drive	61
Samieh Abusaad, Ahmed Benghozzi, Fengshou Gu, Khaldoon Brethee and Andrew Ball	
eMaintenance 2B	69
• Identification of Factors affecting Human performance in Mining Maintenance tasks	71
Mojgan Aalipour and Sarbjeet Singh	
• Asset Management of Aging Process Plant	77
John Andrews and Claudia Fecarotti	
• Issues and Challenges for Condition Assessment: A case study in Mining	85
Esi Maan Nunoo, Phillip Tretten and Aditya Parida	

eMaintenance 3A	95
<ul style="list-style-type: none"> • Computerised Analysis of the Text Entry Fields in Maintenance Work Orders Data 97 Christer Stenström, Mustafa Aljumaili and Aditya Parida • Control Charts supporting Condition-Based Maintenance of Linear Railway Infrastructure Assets 101 Bjarne Bergquist and Peter Söderholm 	
eMaintenance 3B	109
<ul style="list-style-type: none"> • A Conceptual Database Model for Mobile e-Maintenance 111 Jamie Campoos and Erkki Jantunen • Effect of the Shape of Connecting Pipes on the Performance Output of a Closed-loop Hot Water Solar Thermo-siphon 119 Basim Freeegah, Taimoor Asim, Diler Albarzenji, Suman Pradhan and Rakesh Mishra 	
eMaintenance 4A	125
<ul style="list-style-type: none"> • Safety and Security in eMaintenance: The Need for Integration 127 Timothy Tinney, Anders Adlemo, Olov Candell and Ulf Seigerroth • Health and Performances Machine Tool Monitoring Architecture 139 Agustin Prado, Aitor Alzaga, Egoitz Konde, Gabriela Medina-Oliva, Maxime Monnin, Diego Galar, Carl Johansson, Dirk Euhus, Mike Burrows and Carlos Yurre • Procedure for RUL Estimation in Industrial Asset 145 Angel Hernandez, Diego Galar and Numan Perales 	
eMaintenance 4B	151
<ul style="list-style-type: none"> • Multi-Criteria Data Quality Assessment Maintenance perspective 153 Mustafa Aljumaili, Ramin Karim and Phillip Tretten • Big Data Mining in eMaintenance: An Overview 159 Liangwei Zhang and Ramin Karim 	

eMaintenance 1A

Top-of-Rail Friction Measurements of the Swedish Iron Ore Line

Yonas Lemma
Luleå University of Tech.
Sweden
yonas.lemma@ltu.se

Matthias Asplund
Luleå University of Tech.
Sweden
matthias.asplund@ltu.se

Matti Rantatalo
Luleå University of Tech.
Sweden
matti.rantatalo@ltu.se

Jan Lundberg
Luleå University of Tech.
Sweden
Jan.lundberg@ltu.se

ABSTRACT

Friction management in the railway industry is a well-established technology with the aim of optimizing the friction between wheel and rail. Determining the friction coefficient (μ) at the wheel-rail interface is therefore important especially for heavy haul lines with higher axle loads. This paper presents an initial study of the top-of-rail friction condition of a 30 ton axle load, Iron Ore line in the northern part of Sweden. The friction coefficient between the rail and a metal wheel of a portable Tribometer was measured at different geographical locations and during different environmental conditions. The effects of precipitation are studied and compared with the effects of top of rail friction modifiers. The measurements of not lubricated line sections showed values around $\mu \approx 0.6$ compared to $\mu \approx 0.3$ for areas with e.g. top-of-rail lubrication. During snowy conditions a decrease in friction could also be detected.

Keywords

Friction management, Friction measurement, Friction modifier, Heavy haul railway line

1. INTRODUCTION

Determining the coefficient of friction (μ) at the wheel/rail interface is an important diagnostic tool for the freight and transit industry. Application of friction management products to the top-of-rail/wheel tread and lubricant products to the wheel flange/rail gauge interface is critical to ensure long-term benefits, such as increased rail life, reduced lateral forces, and reduced tread/flange wear in addition to reductions in energy consumption and noise levels [1].

Trains operate within the desirable limitations imposed by the friction between the railway wheel and rail surfaces. Inadequate friction causes poor adhesion during braking, which is a safety issue as it exposed to extended stopping distances [2]. Inadequate friction is also a performance issue as it affects traction and thus limits the tangential force that can be developed in curving. Delays occur if a train passes over areas of poor adhesion while in service [2]. Because the tiny contact zone (roughly 1 sq. cm) where steel wheel meets steel rail is an open system, it is exposed to dirt and particles as well as natural lubrication such as humidity, rain and leaves. All of these can seriously affect the contact conditions and the friction forces in the contact. Acceleration and braking usually require a coefficient of friction (the ratio of the tangential load to the normal load) of about 0.2. However, modern power cars and locomotives demand a higher friction coefficient. The friction between the wheels and rail also plays a major role in other wheel-rail interface processes such as rolling contact fatigue (RCF), rail corrugation and noise

generation. If the coefficient of friction is too high, most types of surface damage occur more frequently. Friction coefficients above 0.4 increase the chance of surface fatigue of wheels and rail.

Friction and wear are related, there is no general correlation between the coefficient of friction and normalized wear rates. Tribosystems that have a lower coefficient of friction do not necessarily have a lower specific wear rate. Also, there can be large differences in the specific wear rates of systems that have approximately the same coefficient of friction. [3]. Variations in friction at this interface can cover a wide range of characteristic values and can be dramatically influenced by third-body layer conditions that are dependent on the influence of the environment (e.g. temperature, humidity, precipitation, sunlight) as well as foreign contaminants (either intentionally or unintentionally introduced) including sand, leaf mulch, brake shoe dust, lubricants and friction modifiers [4].

Over the past decades, materials for modifying friction in the wheel/rail interface have been evaluated in laboratories using either a rheometer (pin-on-disk) or the Amsler machine. In the field, the hand-pushed tribometer has been the most effective method of determining μ for dry or conditioned rail/wheel interfaces. More recently, a high-speed production Tribometer (TriboRailer) was developed, adding its own interpretation to the estimates of μ levels in the field. Because these various devices produce somewhat conflicting answers, questions have arisen concerning the accuracy of the 'absolute' measured friction reading both in the laboratory and field [1].

2. FRICTION MANAGEMENT

Friction Management is the process of controlling the frictional properties at the rail/wheel contact to reduce energy consumption, rail wear, lateral forces, environmental issues (noise and corrugations), skid flats, long brake distance, rolling contact fatigue and track structure degradation [1, 5]:

- Lubrication of the gauge face of the rail to minimize friction, wear and curving resistance (μ between 0.1 and 0.25).
- Provide an intermediate friction coefficient (μ between 0.30 and 0.35) at the top of the rail under trailing cars, to control lateral forces in curves and rolling resistance in both curved and tangent track.

There is an expression that what gets measured gets managed, the friction is not an easily observable quantity, heavy haul railways that embark on a mandate for improve friction management need to measure rail and wheel wear, download energy consumption from locomotive and perhaps direct measure friction level with Tribometers to identify any gaps in optimal friction levels [6].

3. FIELD MEASUREMENTS

The Tribometer used in this study was designed and built by the British rail (BR) Research in Derby, England. This tribometer was developed to measure top of rail coefficient of friction in support of braking tests being conducted on new equipment [1]:



Figure 1. Demonstration of hand-pushed Tribometer

The BR Tribometer utilized gravity controlled loading where a standard weight on a lever arm placed a known load through a wheel onto the rail. The wheel was connected to a magnetic clutch so that under normal conditions, the wheel was free to spin the clutch. A manually adjusted variable resistor controlled (reduced) clutch slippage. As slippage was reduced, the resulting force was transferred to an analog weight scale. By increasing resistance of the clutch, longitudinal rolling resistance on the wheel was also increased. The friction at top-of-rail controlled the point at which the wheel would slip. Maximum adhesion was obtained at this point. The scale then showed the force at which wheel slippage occurred.



Figure 2. Schematically diagram of Tribometer

When the operator obtains a steady walking speed, the Tribometer starts a 3 to 5 second measurement sequence. At the end of this sequence, the coefficient of friction of the rail at desired location is displayed on the Tribometer's digital readout on the head. After completing a measurement, the operator may continue to push the unit for additional measurement or stop pushing to manually

record the information. The Tribometer will display the last valid measurement until it is turned off. The wheel speed is determined by measuring the duration of a pulse that is generated by an optical encoder, which is mounted on the support shaft for the measuring wheel. As the wheel speed increase, the duration or period of the pulse decreases. Proper wheel speed is accomplished by pushing the Tribometer along the rail at a normal walking pace (1.5ft/sec or 0.5m/sec minimum). With all initial conditions met, the main board's central processing unit (CPU) will begin a six step test cycle by applying a ramping braking force to the measuring wheel. The braking force is provided by an electromagnetic brake that is attached to the measurement wheels* support shaft. An automatic ramping control circuit immediately senses the point at which wheel slippage occurs and automatically reduces the braking action to the measuring wheel to prevent the wheel from digging into the lubricant on the rail and generating artificially high friction readings [7]. The primary drawback of the BR Tribometer was that it could not be easily redesigned to measure friction on the gauge-face of a rail due to the lever/gravity load mechanism. Since the gauge-face side of rails can wear to a wide range of shapes and slopes, no common wheel angle could be specified. Subsequently, using the BR device would have required a lever system with infinite adjustments in order for a constant load to be maintained. The AAR (Association of American Railroads) embarked on a program to develop a top-of-rail and gauge-face Tribometer. The subsequent AAR prototype utilized the same friction measuring concept as the BR Tribometer. However, instead of a gravity loaded lever, a spring loaded pivot was used to obtain a linear force. Fine tuning of the vertical load or lateral load was controlled by the operator through an adjustment screw.

4. CASE STUDY

The Iron Ore Line (Malmbanan) is a 473 km long track section located in northern Sweden and has been in operation since 1903. This track section stretches through two countries, namely Sweden and Norway, and the main part of the track runs on the Swedish side, where the owner is the Swedish Government and the infrastructure manager is Trafikverket (the Swedish Transport Administration). The ore trains are owned and managed by the freight operator and mining company LKAB [8]. LKAB increased the axle load on Malmbanan line from 25 to 30 t and maximum speed of iron ore train from 50 to 60 km/h. This change is expected to result in higher track geometry degradation levels. In addition to iron ore transportation, the line is used for passenger trains and other freight trains. The passenger train speed from 80-135 km/h. The track consists of UIC 60 rails and concrete sleepers. The ballast type is M1 (crushed granite), and the track gauge is 1435 mm [9].

The Iron Ore line region is subject to harsh climate condition: winter snowfall and extreme temperatures, ranging from -40°C in winter to $+25^{\circ}\text{C}$ in summer [10].

On the selected truck sections, section 119, Sävast and Norra Sunderbyn and section 111, Tornehamn, and Katterjåkk (see Figure 3). The criteria's selected these locations, how is the friction condition on the iron ore line close and relatively far away from the top of rail (TOR) equipment. There is one TOR close to Sävast and two TOR close to Tornehamn iron ore line.

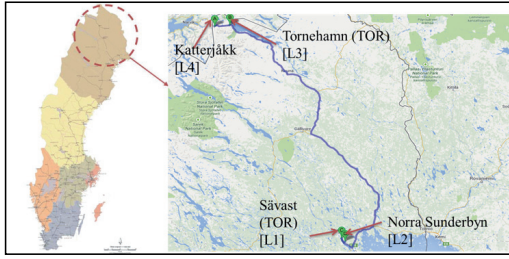


Figure 3. The Iron Ore line located in North Sweden and measurement places, Sävast, Norra Sunderby, Tornehamn and Katterjäck

5. RESULTS AND DISCUSSION

The measurement results are presented in Figures 4 to 10. The first location friction coefficient measurement has taken in Sävast four different dates in different weather condition are shown in figures 4 and 5, left and right rail friction coefficient respectively. During the measurements were performed the rail temperature, humidity and weather conditions are recorded as shown in Table 1. In Sävast there is one TOR lubrication between pole number 9 and 10, however most of the rail line in this location the friction coefficient is higher than the desirable limit. As shown in figures 4 and 5 the friction coefficient is very varied (0.2-0.72) depending on the rail temperature, humidity and other factors. On the final measurement date, (18 March 2014) the friction coefficient is higher than the previous dates of measurements and on this day the humidity was very low 58.5 as shown in Table 1, and the other reasons the friction coefficient is high variation on measurement dates (22 October 2013 and 11 November 2013) compare with last measurement date (18 March 2014) in Sävast was there some contamination on the rail when the first two measurement has taken.

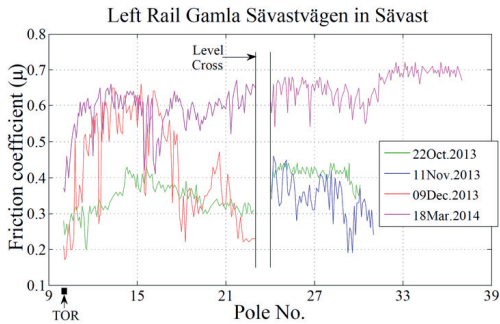


Figure 4. Friction coefficient of left rail Gamla Sävastvägen

The second location measurement has been carried out in Norra Sunderby Knönsvägen approximately 8 km away from TOR equipment. Measurement has taken three times in different dates as shown in Figures 6 and 7 in different weather condition.

Table 1. Summary of statistical friction coefficient measurement in different location

Location	Date of Measurement	TOR	Temperature of Rail in °C	Humidity	Right Rail		Left Rail	
					Mean	SD	Mean	SD
L 1	22 Oct. 2013	Yes	-10	82.5	0.33	0.05	0.37	0.05
	11 Nov. 2013	Yes	-11	90.5	0.39	0.10	0.42	0.10
	09 Dec. 2013	Yes	-22	80.1	0.36	0.12	0.45	0.14
	18 Mar. 2014	yes	-10	58.5	0.67	0.03	0.62	0.07
L 2	22 Oct. 2013	No	-10	79.7	0.59	0.02	0.58	0.02
	11 Nov. 2013	No	-11	81.2	0.47	0.04	0.41	0.03
	18 Mar. 2014	No	-9	45.7	0.67	0.04	0.68	0.03
L 3	16 Oct. 2013	Yes	-2	93.8	0.33	0.16	-	-
	29 Oct. 2013	Yes	-1	91.9	0.55	0.06	0.48	0.13
L 4	29 Oct. 2013	No	-2	93.2	0.52	0.1	0.58	0.08

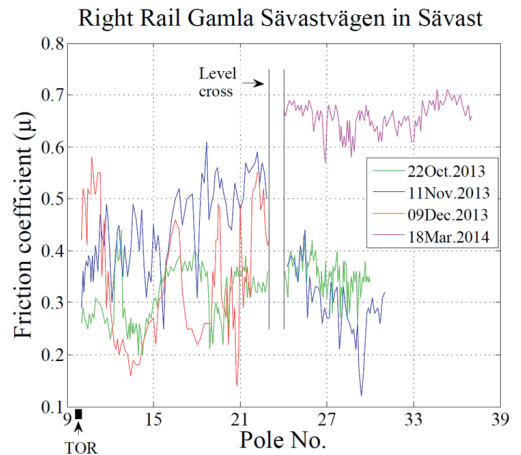


Figure 5. Friction coefficient of right rail Gamla Sävastvägen

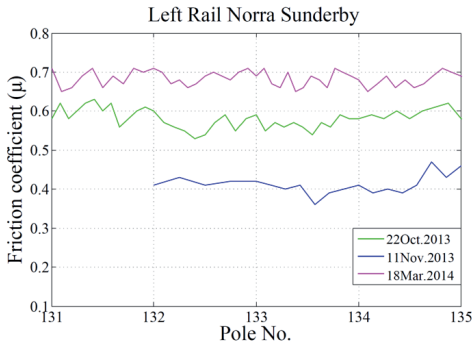


Figure 6. Friction coefficient of left rail Norra Sunderby

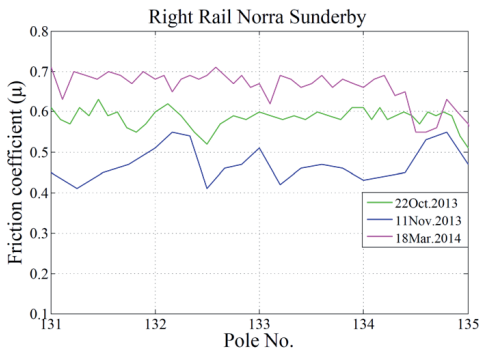


Figure 7. Friction coefficient of right rail Norra Sunderby

The third location friction measurement has taken place in Tornehamn in two different dates, the first day (16 October 2013) is shown in Fig. 8 during the measurement it was snowing, as it can be seen in figure the friction measurements from the tunnel exit (Pole 89) to entrance of snow protection (Pole 103) are under the influence of snow which reduces the coefficient of friction. Inside the tunnel and the snow protection areas the effect of snow does not appear in the measurements and this can be observed clearly in Fig 8.

Another measurement was carried out in the same location after 15 days on (29 October 2014) it was a sunny day, the result as shown in Fig. 9 showed very high friction value recorded even though there are two TOR lubrication in this location near to (Pole No. 68 and 95). As seen in the Figures 8 and 9 in closed area that means inside the tunnel and snow protection the coefficient of friction is very higher than out in air between 0.4 and 0.7 this indicates that the friction value in closed area is higher.

The fourth measurement is carried out in Katterjåkk approximately 20 km far away from TOR equipment. Between pole number 70 and 80 the friction coefficient was not measured. Most of the place along this line the friction coefficient is more than 0.45 as shown in Fig.10 During the measurement between pole number 65 and 75 there was some water contamination

which brought the friction coefficient value to very low value suddenly

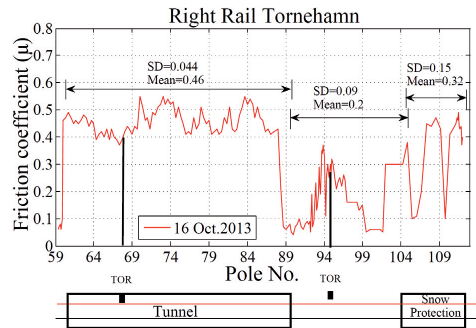


Figure 8. Friction coefficient of right rail Tornehamn

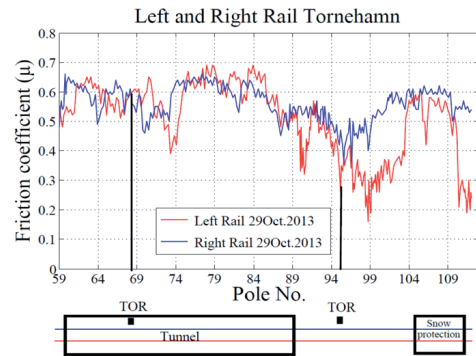


Figure 9. Friction coefficient of left and right rail Tornehamn

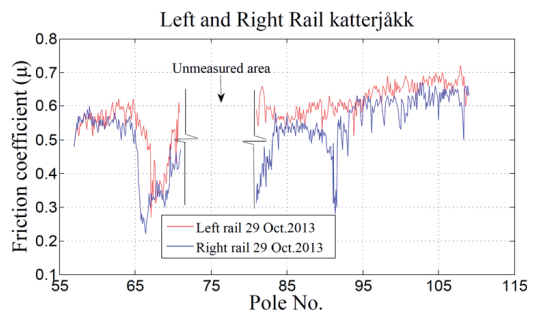


Figure 10. Friction coefficient of right and left rail Katterjåkk

Table 2. Different locations iron ore line annual passing tonnage and FC measured length

Location	Annual passing tonnage (MGT)	Curve radius (M)	Friction coefficient measured length app. (M)
L1	16.54	480	1500
L2	16.54	Tangent	200
L3	27.70	Tangent	2750
L4	27.70	Tangent	2200

KELTRACK TOR lubrication are used in two locations Sävast and Tornehamn, KELTRACK friction modifiers are specifically designed to manage friction levels on the top-of-rail at an intermediate and controlled level of 0.35. Containing no oils or greases, KELTRACK is similar to a latex paint and is designed to dry rapidly in the rail/wheel interface [11]. As shown in Figures 8 and 9 in Tornehamn there are two TOR equipment but the one inside the tunnel seems not properly working, as the results of the friction coefficient measurement value is very high it shows something has problem in TOR.

6. CONCLUSION

As it can be seen in the results of the measured friction, most of the locations have higher friction coefficient value than the desired value of approximately 0.3-0.35. The measurement of the friction coefficient in Tornehamn shows higher values (0.5-0.6) even though there are two TOR equipments in this area.

Measurements performed at Gamla Sävastvägen close to the TOR equipment showed lower friction values compared to the reference point (Norra Sunderbyn Knösavägen), which indicates some form of lubrication.

One interesting conclusion is that the friction measurement performed in Tornehamn shows a distinct difference between the friction coefficients measured in the tunnel, out in the open and under the snow cover. Here, the influence of the open sky can clearly be seen.

At Norra Sunderbyn Knösenvägen and Katterjokk there is no installed TOR lubrication equipment and both of this location show higher friction coefficient values.

Finally, the results show that there is a need for further investigations, which take into account other parameters that significantly influence the friction coefficient (for example roughness, humidity, temperature etc.).

Conclusive remarks for the improved management of the friction in these locations will require some future work and detail research.

ACKNOWLEDGMENTS

The authors are pleased to thank Trafikverket and the technical support of Luleå Railway Research Center for their financial support

7. REFERENCES

- [1] Harrison, H., McCanney, T., and Cotter, J. (2002). Recent developments in coefficient of friction measurements at the rail/wheel interface. *Wear*, 253(1-2), 114–123.
- [2] Iwnicki, Simon, ed. *Handbook of railway vehicle dynamics*. CRC Press, 2006.
- [3] Raymond G. Bayer. (2002). *Wear analysis for Engineering*. New York.
- [4] Oldknow, K., Eadie, D. T., and Stock, R. (2012). The influence of precipitation and friction control agents on forces at the wheel/rail interface in heavy haul railways. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 227(1), 86–93.
- [5] Sroba, P., Oldknow, K., Dashko, R., and Roney, M. (2005). *Canadian Pacific Railway 100 % Effective Friction Management Strategy*.
- [6] IHHA (2009). *Guidelines to best practices for heavy haul railway operations infrastructure construction and maintenance issues*. Virginia, USA.
- [7] LBFOSTER, *Tribometer version 4.02 with dual stabilizer latches user manual*. USA.
- [8] Asplund, M. (2013). *Wayside Condition Monitoring Technologies for Railway Systems*. Licentiate Thesis.
- [9] Arasteh khoy, I., Schunnesson, H., Juntti, U., Nissen, A. and Larsson-Kräik, P-O. (2013). Evaluation of track geometry maintenance for a heavy haul railroad in Sweden - A Case Study. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*.
- [10] Kumar, S., Espling, U., and Kumar, U. (2008). Holistic procedure for rail maintenance in Sweden. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 222(4), 331–344.
- [11] LBFOSTER (2014). *Friction Management*. Available at: <http://www.lbfoster-railtechnologies.com/>, accessed:06 June 2014.

Analysis of gauge widening phenomenon on heavy haul line using measurement data

Stephen M. Famurewa
Luleå Railway Research Centre
Luleå University of Technology
Luleå, Sweden
+46920492375
stefam@ltu.se

Matthias Asplund
Trafikverket
Luleå University of Technology
Luleå, Sweden
+46920491062
matasp@ltu.se

Paul Abrahamsson
NorJeTS
Narvik, Norway
+46703115110
paul.abrahamsson@norjets.no

ABSTRACT

The operational safety and maintenance cost of railway transport are largely influenced by the geometric characteristics of the track. One of the major geometric parameters essential for safe and high performance of railway track is the gauge. The phenomenon of gauge widening is the gradual increase of the gauge size and it is one of the major track failure modes that could cause derailments if it is not effectively restored in due time. Excessive gauge widening is as a result of inadequate lateral track resistance and lateral rail strength capacity as influenced by sleeper, fastener ballast or subgrade conditions. This paper presents an exploratory analysis of gauge measurement data to summarize the main characteristics and explain the progression of gauge widening in different curve types. The measurement data were collected from 2007 to 2013 in curves with different layouts and structures. The growth of gauge over time is analysed and also the variation in the gauge dimension with the layout of the track is presented for engineering considerations. Apparently, gauges of all the curves studied are above the nominal dimension of 1435mm but below the immediate action limit. The result shows that curve radius has significant effect on the rate of gauge widening as it has been shown that tight curves A and B have high rates of deterioration while curves D and E with large radius have relatively low rates of deterioration. A practical application of this study is the use of the presented quality assessment procedure and the estimated gauge widening rate for condition evaluation and maintenance planning.

Keywords

Gauge widening, track geometry, track curvature, data analysis, fastener and sleeper maintenance, maintenance planning

1. INTRODUCTION

Railway transport plays a vital role in industrial logistics as well as mobility of people. Railway heavy haul operation supports large volume industrial activities such as mining, forestry and production. The need and demand for more raw materials and mining products is an issue that requires sustainable, economically efficient and innovative solutions within the railway industry. Better understanding of the behaviour of track and its structural element under loading is required in order to meet the need for higher operational speed, greater freight, reduced time on track for maintenance, higher volume of transport and better

service quality. Advanced knowledge about the behaviour of track will enhance decision making for modernisation of in terms of investment and maintenance activities.

The demand for more railway capacity is a global issue that requires sustainable and efficient solutions. The anticipated increase in traffic volume will see more axles, higher load, greater speed on track, more contact with overhead cable and accelerated usage of infrastructure. To this end, several investigations have been conducted in the field and laboratory to study different load induced displacement phenomenon as well as other degradation processes of the track owing to increased loading condition. Notable among past researches is the field and laboratory experiment that dealt with lateral rail strength for low speed track, as influenced by sleeper and fastener condition to develop criteria and limits for the prevention of excessive gauge widening [1]. The outcome of these experiments gave some useful information about spike pull out strength and sleeper plate vertical and lateral stiffness behaviour [1].

A study of railway derailment conducted in USA between 1998 and 2000 shows that 20% of the reportable derailments incidents were identified as being likely to be influenced by poor wheel-rail interactions. Wide gauge, track alignment, bogie hunting, and wheels with worn tread and flanges were identified to be responsible for 50% of the derailments incidents related to poor wheel-rail interaction. Outstandingly, around 8% of total derailments and approximately 40% of all derailments related to poor wheel – rail interaction were reported to be caused by wide gauge [2]. This reports calls for special attention on the unique challenge of ensuring that wheels stay on the rail and do not fall off or into the track due to gauge widening or similar phenomenon. Gauge widening resulting from loss of adequate rail or fastening restraint has been identified in other studies as one of the major track failure modes in railways that have caused a large number of derailments [3].

Furthermore, researches and advancement on track stability and geometry quality has led to the suggestion of different intervention levels based on best practices. In order to ensure safe operation of trains, optimise vehicle ride quality, dynamic loading of the track and track geometry maintenance works, the European Standard [4] has set out quality levels and defined a minimum track geometry quality. Three basic intervention levels namely: alert limit, intervention level and immediate action limit have been suggested in standards and guidelines for best practices [4],

[5], [6]. The recommended values of these limits are usually given as a function of speed and are useful for assessment and analysis geometry parameter including gauge to ensure safe and high performance of the track.

In connection with the need to improve railway performance and operation for heavy and long hauls of iron ore, a major investment on infrastructure was initiated in 2006 basically in the northern section of Swedish iron ore line from Riksgränsen to Kaisepakte. The investment was a complete track renewal that included the change of the rail from BV50 to UIC60 and also Hard Wood sleepers (beach) to concrete sleeper. The fastening system was changed from heavy-back system to Pandrol e-clip and Fast Clip while fastening systems in S&C were replaced with Vossloh fastenings. In addition to the investment on infrastructure done by the infrastructure manager, the major freight operator LKAB also replaced the old DM3-locomotives and UAD-wagons to Iore-locomotives and Fanno-wagons respectively (see Figure 1 for the Iore-locomotive). The investments on and upgrading of both infrastructure and rolling stocks were crucial for the increment of railway operation in terms of heavier haul of 30 tonnage axle loads and longer haul of 750 m on the northern loop of the iron ore line. This investment was succeeded by measurement campaigns which commenced in 2007 to understand the behaviour of different track structure under loading in different layout for improved maintenance practices. One of the necessities of these measurements is to further investigate and describe gauge widening and rail restraint characteristics of the track from field measurement point of view after increasing the traffic load.



Figure 1. Two locomotives that pull 68 wagons with a gross weight of 8500 tonnage

The contribution of this paper is the analysis of track gauge measurement data to provide engineering insight into the progression of a typical lateral track displacement. It also gives an assessment of different curves in different sections on a heavy haul corridor as required for maintenance planning and also to determine the progression rate of this phenomenon in different track layout.

The organisation of the paper is as follows: Section 2 provides a brief highlight of track geometry parameters with emphasis on track gauge and other types of lateral displacement phenomenon. Section 3 describes the study area and measurement procedures. Section 4 presents the result of the study and analysis to explain gauge widening phenomenon. Discussion is given in Sections 5

with useful information to support maintenance and future design considerations. The final section presents the findings, suggestions and concluding remarks of the study.

2. TRACK GAUGE

The geometry of track is a measure of its integrity and quality, and it has been adequately proven to be a necessary and not luxurious requirement in the design, construction, installation and maintenance of track [7]. The geometry and irregularity of ballasted tracks are described using principal parameters such as longitudinal level, alignment, cross level, twist and gauge [8], [9].

Track gauge as shown in Figure 2 can be described as the smallest distance between lines perpendicular to the running surface intersecting each rail head profile at point P within a range from 0 to Z_p (14mm) below the running surface [8].

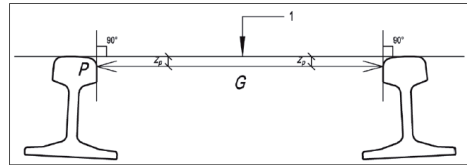


Figure 2. Measurement of track gauge [6], [8]

The stability of track is maintained by the lateral restraint or confining pressure from the ballast, sleeper, fastening system, subgrade or the entire track structure [10]. High lateral force generated at the contact between the wheel flange and the rail due to poor operation pattern or poor design profiles can cause different accidental phenomenon such as wheel climb, gauge widening or rail rollover [2], [11]. On the other hand deteriorating condition of the track structure, fouling of the ballast and poor subgrade condition contribute significantly to the instability of the track and possibly gauge widening. The deterioration contributes to the reduction of lateral guidance provided by rail and consequently increases the risk of derailment. High lateral force generated by the contact between the wheel flange and the rail or low lateral restraint of the track system can cause lateral rail displacement type known as gauge widening which can lead to a wheel/rail separation as shown in Figure 3 [11]. It has been established that high lateral force is common in curves and it is usually induced by a large angle-of-attack of the wheelset [2]. Also, a study on heavy haul curves has shown that there is interaction between gauge widening and other deterioration mechanisms and damage such as rail wear, rolling contact fatigue and fastener distortion [12].

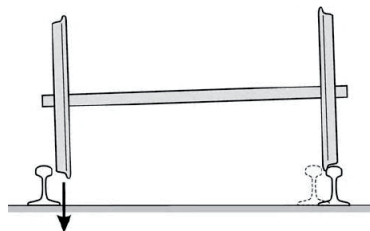


Figure 3. Gauge widening [11]

In practice, intervention limits are recommended by infrastructure manager for monitoring and assessing gauge to ensure safe and dependable railway operation. For example, the limits used by Trafikverket for the Swedish railways are shown in Table 1.

In addition to gauge widening, track panel shift is another related instability phenomenon. Track panel shift is a derailment mechanism resulting from accumulated lateral displacement of the track structure including rails, baseplates, sleepers, fasteners over the ballast and it is basically caused by repeated lateral axle loads [2], [11]. Track with poor initial quality or newly laid or maintained track without adequate consolidation are often characterized with low lateral track strength and stiffness that is subjected to low resistance to lateral force. Soft subgrade also contributes to the lateral movement of the track over the ballast. The occurrence and progression of this phenomenon is accelerated with increase in speed and load with continuous welded rail in curves and poorly aligned tangent. The manner of acceleration and braking has also been observed to induce large lateral forces that can cause panel shift and other lateral displacement phenomenon [2].

Table 1. Limits for track gauge deviation from nominal gauge (1435mm) - for isolated defects [6]

Speed (km/h)	AL		IL		IAL	
	Min	Max	Min	Max	Min	Max
$V \leq 40$	-4	15	-4	22	-10	35
$40 < V \leq 80$	-4	15	-4	22	-10	35
$80 < V \leq 120$	-4	12	-4	17	-9	33
$120 < V \leq 160$	-3	12	-4	15	-8	33
$160 < V \leq 200$	-3	10	-4	12	-7	28
$200 < V \leq 250$	-3	8	-3	10	-5	28

AL is alert limit that if exceeded requires that the track geometry condition is analysed and considered in the regularly planned maintenance operations. IL is the intervention limit that if exceeded requires corrective maintenance in order that the immediate action limit shall not be reached before the next inspection. IAL is the immediate action limit that if exceeded requires that instant measures should be taken to reduce the risk of derailment to an acceptable level.

3. DATA COLLECTION AND ANALYSIS

In connection to the objective of the project described earlier and to study gauge widening phenomenon, the study area with the selected sections and curves are described below. Since the analysis is from field measurement point of view, the measurement and data collection procedure is described afterwards.

3.1 Description of study area

The study area in this article is the northern section of the iron ore line that is the only heavy haul line in Sweden and Europe. The track section is a single track with a length of 127 km, axle load of 30 metric tonnes and annual load of about 30 MGT. The track section is one of the busiest sections in Sweden and has the largest predicted traffic increase of about 136% between 2006 and 2050

due to expansion of the mining industry in the north of Sweden [13]. The track section has a mixed traffic with high range of train speeds and loads including passenger, iron ore and other goods traffic.

In the project, five types of curves in six different sections on the iron ore line were chosen; the locations of the sections are shown in Figure 4. Due to the amount of data available and to avoid inconsistency in the analysis, sections 1 and 2 were left out in this study while sections 3 to 6 were analysed.



Figure 4. Study area on the Iron ore line with section 1-6

The station area, layout and characteristics of the sections are described in Table 2. The length of the sections varies from 1511 m to 6582 m with at least a curve. Further description of the curves on the sections is given in

Table 3. The layout of the sections is categorized into 6, five curves with different radii and a tangent track. The tight curves with radius less than 550 m are categorized as “A curves” while those with radius greater than 850 m as “E curves”.

Table 2. Description of the sections on the study area

Section	Station area	Curve Categories	Number of Curves	Total length (m)
3	Stordalen-Abisko	B, C, E	4	2378
4	Björkliden-Kopparåsen	A, B, C, D, E	18	6582
5	Läktatjåkka-Vassijaure	A, E	3	1511
6	Vassijaure-Riksgränsen	A, B	9	5209

Table 3. Different categories of curves and tangent track

Category	Layout	Radius (m)	Count
A	Curve	<550	9
B	Curve	550-650	17
C	Curve	650-750	3
D	Curve	750-850	2
E	Curve	>850	6
T	Tangent track		6

3.2 Measurement procedures

The measurement of the track gauge is done with handheld tool called MiniProf. The MiniProf measurement is manually done by one man and the system is capable of measuring the cross sectional profiles of the rail in addition to the track gauge. The measurement accuracy of MiniProf for rail profiles is $\pm 9 \mu\text{m}$ [14]. Figure 5 shows the field measurements done with MiniProf on the southern part of the Iron Ore Line.



Figure 5. Field measurements with MiniProf on the southern part of the Iron Ore Line.

The poles of the overhead line are used as markers in the different sections to ensure that gauge measurements are done at the same point for several measurement runs. Basically, two measurements campaign are carried out in a year from 2007 to 2013. This is

done between the June and October. The first measurement is carried out before the annual rail grinding while the other is done after the grinding operation. In total there are 12 measurements at the same points in the different curves in all the track sections that are studied.

4. RESULT AND DISCUSSION

The data collected is analysed to have knowledge about the position of the track and the amount by which they move especially the track gauge as a result of increased loading. The analysis carried out include comparison of gauge characteristic at different track layout, quality assessment of the last measurement data and estimation of the gauge widening rate for different curve categories.

4.1 Gauge at different layouts

The measurements were done in curves, transition curves and tangent segments adjacent to the curves. Figure 6 shows that mean gauges are consistently higher in measurement points on circular curves than transition curves. Furthermore, gauges measured at the transition curves are higher than the tangent segments for all the sections and measurement runs. This is due to unbalance forces in circular curves and also high magnitude of lateral force at the wheel and rail interface that pushes the rails outward in curves against the lateral restraint of the track structure. Greater load and higher line speed have been reported to intensify the lateral force in both circular curves and poorly aligned tracks. In the case of this study area, one would easily relate the distinct higher gauge in circular curves to high loading condition.

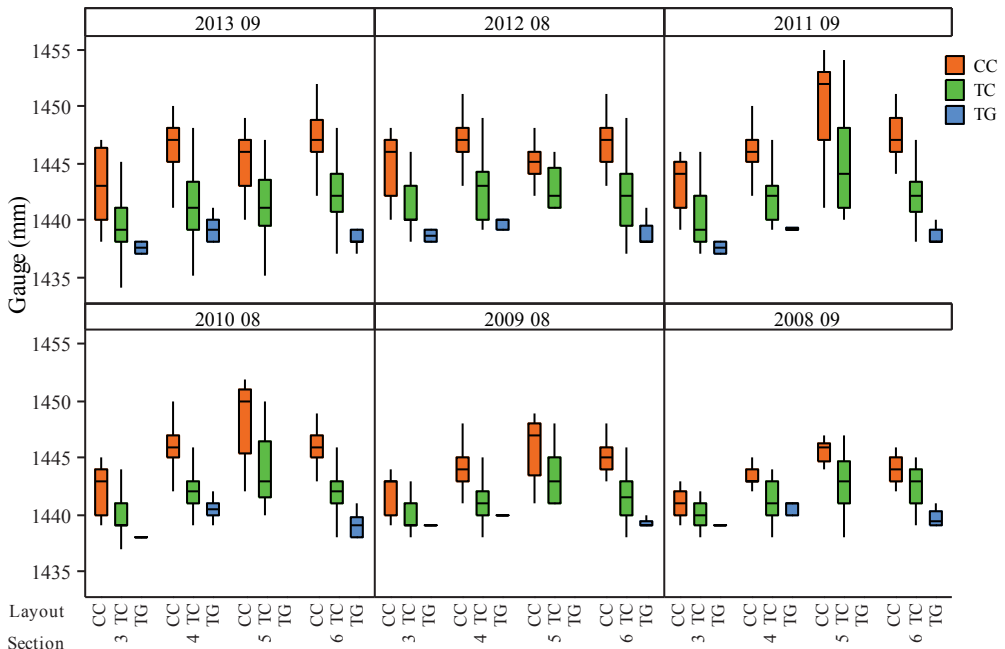


Figure 6. Gauges at circular curve (CC) transition curves (TC) and tangent segments (TG) for the four sections

4.2 Gauge assessment using maintenance limits

The track quality assessment procedure using cumulative frequency distribution plot suggested in the standard [15] is adapted for the analysis of the measured gauge data. The procedure gives an overview of the condition of the track gauge for each section during the last measurement campaign in September 2013. The recommended limit for track gauge deviation from nominal is used to analyse the gauge quality of the sections. It can be seen from Figure 7 that approximately 98% of the measurements in all the sections are above the nominal gauge. However, the distribution differs at higher gauge especially for section 4, 5 and 6 with obvious deviation from nominal gauge. A reason could be that these three sections have the type-A curve which has small radius and the magnitude of the lateral force is most likely to be high. Relating this plot with the recommended maintenance limit– it can be seen that 25%, 23% and 14% of the measurement points are above the alert limit in sections 6, 4 and 5 respectively while in section 3 no point is beyond the alert limit. A reasonable explanation for this is the composition of the curve types in the different sections besides the geotechnical condition and modulus of the fasteners. Section 6 happens to be peculiar in the quality chart due to the fact that it has five type-A curves and two type-B curves which are basically tight curves. From maintenance planning perspective, this quality description is an indication of track condition and need for intervention.

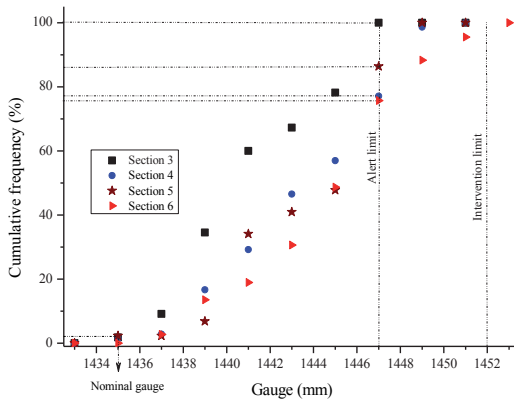


Figure 7. Quality assessment of track gauge using cumulative distribution plot

4.3 Gauge widening

The progression of gauge widening is investigated herein to improve the safety and operational performance of the line. The measurement data collected from 2008 to 2013 are plotted to assess the effect of the increase in axle load as against the major investment made on both track and freight vehicles. Figure 8 shows the deterioration of gauge of the different circular curves in sections 4-6. The initial gauge of circular curve A, B and C are apparently high due to their small radius and relatively weak fastening system. For curves D, E especially in section 4, their initial gauges are close to the nominal dimension and the rate of their deterioration is low. The slope of the plots shown in Figure 8 is a measure of the deterioration rate of gauges at each curve and section. It can be inferred from the slope of the plots that the deterioration rate of curves A, B and C are higher than that of curves D and E. Furthermore, the plot shows there is improvement in the two curves in section 5 in the year 2012 due to replacement of Pandrol e-clip with a stronger fastening design, Pandrol e+.

An interesting issue from the infrastructure manager's point of view is to estimate the rate of widening of the gauge or the deviation from nominal dimension at the different sections and curves. The widening rate is one of the numerous measures for assessing the impact of axle load increase on line as against the upgrading of the line. Table 4 gives the estimate of the widening rate typical for the curves in the different sections using simple linear regression method and least squares model as the fitting procedure. The goodness of fit of the linear regression is confirmed using the co-efficient of determination (r-square); the r-square values for almost all the fits are greater than 0.9. The table shows that the widening rate for Type A and B curves are over 1mm/year while that of Type D and E curves are less than 1mm/year. Type C curves are eventually expanding at a relatively high rate as the estimated rate ranges from 0.81 to 1.30. The variation of the degradation rate for the same curve in different section is an indication that other factors are also significant for the high pull out force in circular curves or low lateral restraint of the track. Notable among these factors are geotechnical condition, speed at the curves, geometry condition, wheel-rail contact condition, bogie features, modulus of the fastener, and operation characteristics such as aggressive braking.

It is essential to emphasise that for a reliable prediction of gauge condition at the investigated site, simple linear regression approach can be used with the estimated deterioration rate and appropriate error correction technique. However, in case of higher prediction accuracy, it is better to use techniques such as: grey model, neural network, support vector machine and other statistical non-linear regression models.

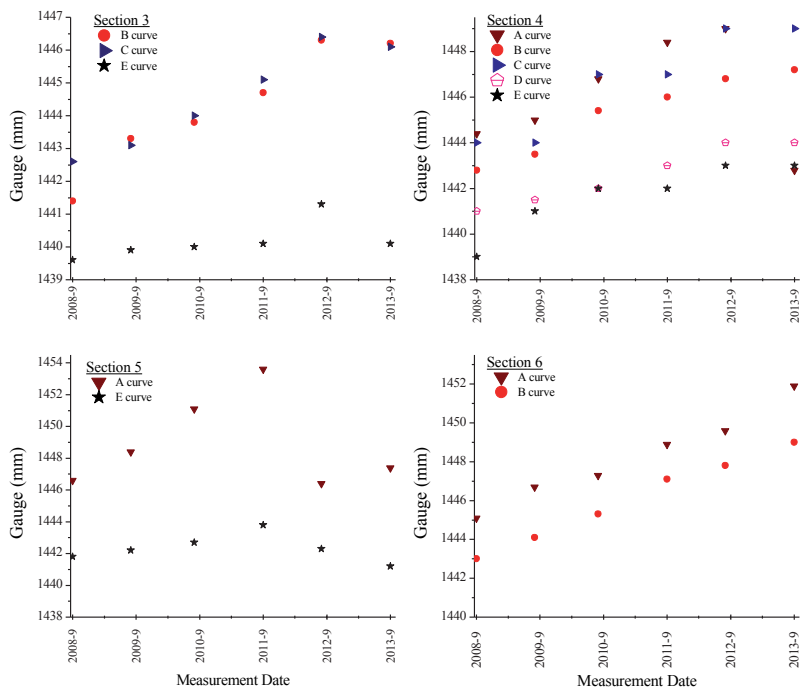


Figure 8. Progression of gauge widening at the four sections and five curve types

Table 4. Typical degradation rate for curve types at the different sections

Curve	Degradation rate (mm/year)				
	Section 3	Section 4	Section 5	Section 6	Average
A		1.26	2.37	1.27	1.6
B	1.12	1.05		1.23	1.1
C	0.81	1.30			1.1
D		0.75			0.8
E	0.36	0.90	0.65		0.6

5. CONCLUSION

This paper has contributed to an important issue related to safety and performance of railway track. The actual focus of the paper is the analysis of track gauge measurement data to provide engineering insight into the progression of a typical lateral track displacement. Also, an assessment has been done for five curve types in four sections on a heavy haul corridor as required for maintenance planning. The concluding remarks include:

- The quality assessment procedure presented in this paper shows that 25%, 23% and 14% of the measurement points during the last measurement run are above the alert limit in

sections 6, 4 and 5 respectively while none is above this limit in section 3.

- Circular curves have been established to have wider gauge in comparison to adjacent transition curves and tangent segments owing to the high magnitude of lateral force at the wheel and rail interface that pushes the rails outward in curves.
- The mean gauge widening rate is between 0.6 and 1.6 mm/year depending on the size of the curve. Curve radius has significant effect on the rate of gauge widening as it has been shown that tight curves A and B has a high rate of deterioration while D and E have relatively low rate of deterioration.

- Other factors also contribute to the rate of deterioration of track gauge besides the curve radius.

In the future, the study will be extended to investigate the correlation between gauges and wear rate of rail and also occurrence of RCF and other failure mode such as fastener damage.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support of Trafikverket, SPENO and Luleå Railway Research Centre (JVTC). The support of Mr Per Gustafsson of LKAB during the measurements is acknowledged.

REFERENCES

- [1] A. Kish, D. Jeong and D. Dzwonczyk, Experimental Investigation of Gauge Widening and Rail Restraint Characteristics 1984.
- [2] S. Iwnicki, Handbook of railway vehicle dynamics, CRC Press, 2006.
- [3] C. Ulianov, "Holistic overview of freight train derailments in Europe: causal and impact analysis," in 10th World congress on railway research, 2013.
- [4] European Committee for Standardization (CEN), "EN-13848-5; railway applications – track – track geometry quality –Part 5: Geometric quality levels – plain line," European Standard 2010.
- [5] UIC, "Best practice guide for optimum track geometry durability," 2008.
- [6] Trafikverket, "Track structure – Track quality – Requirement during construction and maintenance (Banöverbyggnad – Spårläge - krav vid byggande och underhåll)," Trafikverket, Börlange, Sweden 2014.
- [7] S.M. Famurewa, T. Xin, M. Rantatalo and U. Kumar, "Comparative study of track geometry quality prediction models," in 10th international conference on condition monitoring and machinery failure prevention technologies, 2013.
- [8] European Committee for Standardization (CEN), "EN-13848 - 1; railway applications – track – track geometry quality – part 1: Characterisation of track geometry," European Standard 2008.
- [9] B. Lichtberger, Track Compendium: Formation, Permanent Way, Maintenance, Economics. Hamburg, Germany, Eurailpress, 2005.
- [10] B. Indraratna, W. Salim and C. Rujikiatkamjorn, "Advanced rail geotechnology-ballasted track," 2011.
- [11] A.A. Shabana, K.E. Zaazaa and H. Sugiyama, Railroad vehicle dynamics: a computational approach, CRC Press, 2010.
- [12] P. Abrahamsson and D. Rhodes, "A study of interacting modes of track deterioration in heavy haul curves," in 10th World congress on railway research, 2013.
- [13] Trafikverket. "Prognosis of the Swedish goods' flow for the year 2050 (Prognos över svenska godsströmmar år 2050: underlagsrapport. Publikationsnummer: 2012:112)", Trafikverket, Börlange, Sweden 2012.
- [14] Greenwood Engineering. MiniProf-Rail [Online]. available: <http://www.greenwood.dk/miniprofrail.php>.
- [15] prEN-13848-6, "Railway applications - track - track geometry quality - part 6: Characterisation of track geometry quality," European Standard 2012.

Context Awareness And Railway Maintenance

Roberto Villarejo
Luleå Tekniska Universitet
+46 920491000
roberto.villarejo@ltu.se

Diego Galar
Luleå Tekniska Universitet
+46 920491000
diego.galar@ltu.se

Carl – Anders Johansson
Luleå Tekniska Universitet
+46 920491000
carl-anders.johansson@ltu.se

Manuel Menendez
Vias y Construcciones S.A.
+34 915969747
manuel.menendez@vias.es

Numan Perales
Luleå Tekniska Universitet
+46 920491000
numan.perales@ltu.se

ABSTRACT

A railway is an extremely complex system requiring maintenance decision support systems to gather data from many disparate sources. These sources include traditional maintenance information like condition monitoring or work records, as well as traffic information, given the criticality of maintenance in avoiding traffic disruptions and the need to minimise the track possession time for maintenance.

A methodology is required if maintainers are to understand the data as a whole. Context engines try to link the various data constellations and to define interactions within the railway system. This is not easy since data have different natures, origins and granularity. But if all information surrounding the railway asset can be considered, decisions will be more accurate and problems like false alarms or outlying anomalies will be detected. The contextualisation of the data seems to be a feasible way to allow condition monitoring data i.e physical measurements and other variables, to be understood under certain conditions (weather, regulations etc.) and as a consequence of certain actions (maintenance interventions, overhauls, outsourcing warranties etc.).

This paper proposes the use of context engines to provide meaningful information out of the overwhelming amount of collected and recorded data so that proper maintenance decisions can be made. In this scenario, fluffy information coming from work orders and expertise of maintainers is a big issue since such information must be converted to numerical values. The fuzzy logic approach seems a promising way to integrate such information sources for diagnosis.

Keywords

Context driven, Railway maintenance, Maintenance Knowledge Management, context awareness, contextual decision making.

1. RAILWAY COMPLEXITY AND THE ROLE OF MAINTENANCE

Rail transport will play an important role in the future if capacity can be increased. If this is to be accomplished, it is necessary to improve the competitiveness of railways by ensuring a sustainable, efficient and safe service.

In the last 20 years, the European Commission has worked to restructure the European rail transport market and strengthen the position of railways with respect to other transport modes. The Commission's efforts have concentrated on three major areas, all crucial for developing a strong, transparent and competitive rail transport industry:

- Opening rail transport to market competition.
- Improving the interoperability and safety of national networks.
- Developing rail transport infrastructure.

If these goals are to be achieved, it is essential to improve the interoperability and safety of national networks and promote a single European rail market. There are several barriers to overcome, however, as there is no common definition of standards at the European level.

The railway business is large, integrated, automated and complex; providing safe, reliable and punctual service has become a strategic issue for railway administrators as they seek to meet customer requirements and gain a competitive advantage.

Over the past decades, there have been significant improvements (mainly on high speed lines) in railway safety and services but passengers still expect affordable and on-time service.

We can say that operators + railway administrators = railway networks. These components must work together in harmony. It is essential to understand the operation of trains and their role in this complex relationship. Indeed, much of the day-to-day work of railway professionals should be oriented toward relations with the other components of the railway network, components outside their direct control.

There are two main stakeholders in the railway. On one hand, operators for passengers or freight traffic own or rent the trains and buy time slots and services from the railway administrators. On the other hand, railway administrators own the infrastructure and provide logistic support for operation and maintenance (preventive or corrective actions).

Both stakeholders produce huge amount of data coming from operation and maintenance actions, creating databases that should

be able to improve future O&M activities. There is major concern about how these data can be used and many tools very popular in ICT disciplines have recently been imported into the railway sector.

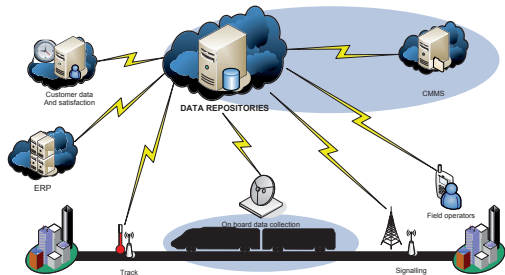


Figure 1. Information data sources to be considered in railway data mining

Data mining plays a key role in extracting, organising, and analysing large sets of data to analyse and draw meaning from them. Data are what we collect and store, and knowledge helps us make informed decisions. The extraction of knowledge and information from data is called data mining. Data mining can also be defined as the exploration and analysis of large sets of data in order to discover meaningful patterns and rules. The first step is to find consistent patterns and/or systematic relationships between variables and the second is to validate the findings by applying the detected patterns to new subsets of data.

Railway operators and managers are mining more and more data collected from trackside and handheld readers, onboard locomotive devices and integrated systems for an array of purposes; see Figure 1. The challenge for users is sorting these new-found data, interpreting them and using them to get better at keeping freight moving. For technology providers, the challenge is to keep abreast of the needs of their increasingly diverse customer bases.

2. DISPARATE MAINTENANCE DATA SOURCES FOR RAILWAY HEALTH ASSESSMENT

Railway infrastructure, such as railroads, has a direct impact on the shutdown or slowdown of railway vehicles. The condition and maintenance of these assets is critical to the effectiveness, efficiency and security of a train. Any improvement in the condition or maintenance management of linear assets and the technology involved in maintenance tasks can have a substantial influence on the operation of the corresponding rolling stock using this railroad.

Therefore, there is a need to integrate railroad information and rolling stock to get an accurate health assessment of the vehicles and determine the probability of a shutdown or slowdown [1]. For railroads and rolling stock, much information needs to be captured and analysed to assess the overall condition of the whole system, i.e. infrastructure plus vehicles.

Examples of information that can be collected include track availability, use of track time, track condition, performance

history, and the work performed. Measurements of the condition of a linear asset such as track typically include continuous and spot measurements from automatic inspection vehicles, visual inspections from daily walking inspections, and records of in-services failures. Examples of conditions measured by automatic inspection vehicles include geometry car measurements (deviation from design curves, geometry exceptions to railroad standards, vehicle ride quality exceptions), rail measurements, gage restraint measurements, track deflection and stiffness measurements, clearance measurements, and substructure measurements.

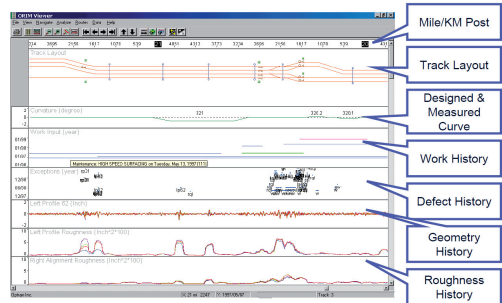


Figure 2. Information data sources to be considered in railway data mining

The information about a railroad is usually collected and maintained, for example, in a set of track charts or line books (see Figure 2). A track chart is the linear representation of all infrastructure assets along a linear asset based on a marker posts and offset measurement system. Updating the track charts generally occurs on an ad hoc basis, so discrepancies, missing facilities, and incorrect location information are common. Even with a complete and accurate map of the corridor, the rail, ties, and other corridor assets do not have any physical characteristics that lead to easy identification. Furthermore, problem areas for targeted maintenance often do not obey discrete physical boundaries such as the beginning and end of a rail section.

The development of a variety of track condition indicators such as geometry cars, rail defect detection equipment and gage restraint management systems have resulted in a significant amount of new and useful information for track maintenance. But a large amount of information provided over a large area quickly leads to information overload.

Moreover, much of the data collected about tracks and rolling stocks is dispersed across independent systems that are difficult to access and are not correlated. If the data from each of these independent systems are combined into a common correlated data system, this system could provide a rich new set of information that greatly adds to the value of the individual systems. For example, it is common for facilities like railroads to collect work records of where work has been done. Railroads also typically measure the quality of their tracks to see where work needs to be done. However, these two data sets remain in separate and individual systems. By combining the data into a location correlated dataset, i.e. metadata (Figure 2), the quality and/or the effectiveness of the work being performed can be analysed by comparing the track quality before and after the work at the location where the work was completed.

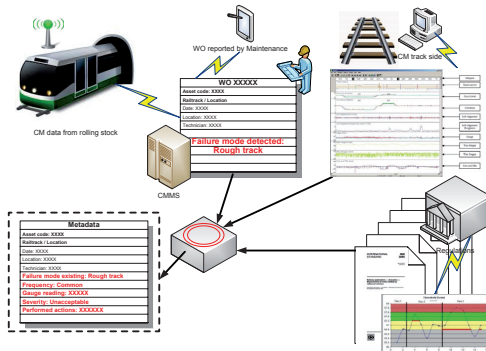


Figure 3. Metadata for maintenance railway knowledge extraction

The greatest challenge for improving asset performance is that the necessary information is scattered across disconnected silos of data in each department. It is difficult to integrate these silos due to their fundamental differences. For example, control system data are real-time data measured in terms of seconds, whereas maintenance cycle data are generally measured in terms of calendar based maintenance (e.g., days, weeks, months, quarters, semi-annual, annual), and financial cycle data are measured in terms of fiscal periods.

CMMS (work orders) and CM (track geometry and other physical measurements) are the most popular repositories of information in railway maintenance, and they include data on most of the deployed technology. Unfortunately, isolated information islands are often created. While using a good version of either technology can lead to the achievement of maintenance goals, combining the two into one seamless system can have exponentially more positive effects on a maintenance group's performance than either system alone might achieve. The strengths of a top-notch CMMS (preventive maintenance (PM) scheduling, automatic work order generation, maintenance inventory control, and data integrity) can be combined with the wizardry of a leading-edge CM system (multiple-method condition monitoring, trend tracking, and expert system diagnoses) in such a way that work orders are generated automatically based on information provided by CM diagnostic and prognostic capabilities. Just a few years ago, linking CMMS and CM technology was mostly a vision easily dismissed as either infeasible or too expensive and difficult to warrant much investigation. Now, the available technology in CMMS and CM has made it possible to create a link relatively easily and inexpensively.

A top-notch CMMS can perform a wide variety of functions to improve maintenance performance. It is the central organisational tool for World-Class Maintenance (WCM). Among other critical features, a CMMS is designed to facilitate a shift in emphasis from reactive to preventive maintenance. It achieves this shift by allowing maintenance professional to set up automatic PM work order generation. A CMMS can also provide historical information which is then used to adjust PM system setup over time to minimise unnecessary repairs, while avoiding run-to-

failure repairs. PMs for a given piece of equipment can be set up on a calendar schedule or a usage schedule linked to meter readings. A fully-featured CMMS includes inventory tracking, workforce management, and purchasing, in a package that stresses database integrity to safeguard vital information. The result is optimised equipment up-time, lower maintenance costs, and better overall efficiency.

For its part, a CM system should accurately monitor real-time equipment performance and alert maintenance professional to any changes in performance trends. There are a variety of measurements that a CM package might be able to track, including track geometry, lubrication condition, temperature, ultrasound inspection etc. These measurements are captured by monitoring tools like ferrographic wear particle analysis, proximity probes, triaxial vibration sensors, accelerometers, lasers, and multichannel spectrum analysers. The very best CM systems are expert systems that can analyse such measurements like vibration and diagnose machine faults. Expert system analysis like this puts maintenance procedures on hold until absolutely necessary, thus ensuring maximum equipment up-time. In addition, the best expert systems offer diagnostic fault trending where individual asset fault severity can be observed over time.

Both CMMS and CM systems have strong suits that make them indispensable to maintenance operation improvements. CMMS is a great organisational tool, but cannot directly monitor equipment conditions. A CM system excels at monitoring those equipment conditions, but is not suited to organising the overall maintenance operation. The logical conclusion, then, is to combine the two technologies into a single system that avoids catastrophic breakdowns but eliminates needless repairs to equipment that is running satisfactorily.

The general opinion among maintenance staff is that the application of information technology brings dramatic results in machine reliability and maintenance process efficiency. However, few maintenance managers can show or calculate the benefits of the application of information technologies. Technology providers are trying to develop increasingly advanced tools while maintenance departments struggle with the daily problems of implementing, integrating and operating such systems. The technology providers or the users generally do not know the feasibility of applying these technologies; they only know that they seem to improve the efficiency of the maintenance activities. The users combine their experience and heuristics to define maintenance policies and employ condition monitoring systems. The resulting maintenance systems are a heterogeneous combination of methods and systems in which the integrating factor is the maintenance personnel. The information about maintenance is stored in these human minds, forming an organisational information system and creating a high reliance on the expertise of the maintenance staff.

With the emergence of intelligent sensors to measure and monitor the health state of a component and the gradual implementation of information and communication technologies (ICT) in organisations, conceptualisation and implementation of e-maintenance is becoming a reality [8]. While e-maintenance shows promise, seamless integration of information and communication technologies (ICT) into the railway environment remains a challenge. It is critical to understand and address the requirements and constraints from the maintenance as well as the ICT standpoints in tandem.

3. CONTEXT AWARENESS AND RAILWAY MAINTENANCE

3.1 Contextual Information

Words are a basic form of data in social science research because they are the usual medium of social exchange. In many cases, insight into meanings can be obtained by examining profiles of ideas and contextual information contained in text. By "text" we mean a transcript of naturally occurring verbal material, including conversations, written documents such as diaries or organisation reports, books, written or taped responses to open-ended questions, media recordings, and verbal descriptions of observations. In sum, the transcript is a computerised file of conventional words and sentences for one or more cases.

Clearly, methodologies for directly, systematically and efficiently handling textual data are needed. Trained coders are commonly used but serious validity, reliability, and practical problems are often encountered.

A context-aware system actively and autonomously adapts and provides the most appropriate services or information to users, taking advantage of people's contextual information while requiring little interaction.

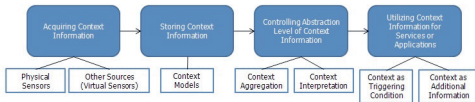


Figure 4. A General Process in Context-Aware Systems

Context-aware systems are usually complicated. They are responsible for many jobs, including representation, management, reasoning, and analysis of context information, and require the collaboration of many different components in the systems. Because there are many types of context-aware systems, it is hard to generalise but a context-aware system usually follows four steps.

The first step is to acquire various types of context-aware information from sensors converting real world context information into computable context data. The system stores these data in its repository. The kind of data model used to represent context information is very important. Context models are diverse, and each context model has its own unique characteristics. To easily use the stored context data, the system controls their abstraction level by interpreting or aggregating context data. Finally, the system uses the abstracted context data in context-aware applications.

In the following sections, we explain how a general context-aware system works and discuss related issues. Figure 4 shows a general process in a context-aware system.

3.2 Protocol To Destroy Communication Barriers

Working within the context defined by ISA-95 ensures this information can be used by higher level enterprise applications like ERP or EAM. The emerging standard is specifically focused on providing value to end users by creating plug and play capabilities for faster implementation and by allowing users to pick and choose the best solutions from complying suppliers. An

extensible, open architecture based on XML and service oriented interfaces that leverage best of breed technology and support practical interoperability and compliance are implicit in Open O&M.

XML is arguably the most popular protocol for the communication exchange of maintenance information [12]. While HTML is focused on document format, XML considers information content and relationships. A class of software solutions is evolving which enables tighter coupling of distributed applications and hides some of the inherent complexities of distributed software systems. The general term for these software solutions is middleware. Fundamentally, middleware allows application programs to communicate with remote application programs as if the two were located on the same computer. XML use in communications is discussed in [2], [3], and [6].

The process to transfer information between disparate sources in an XML environment is as follows. Data from each of the computers involved in asset data exchange must be wrapped in an XML wrapper and sent to an XML data server. Because XML is a descriptive language, the server can process any type of data. At the server, if necessary, the data are encapsulated and mapped to a new XML wrapper, i.e., these data are mapped from one XML schema to one or more XML schemas created for each of the receiving applications.

An XML model describes the structure of information. XML schemas can be used to test the validity of documents; this is especially important when web-based applications are receiving and sending information to and from many sources. When mapping the process model to an XML schema, certain rules need to be established; there are multiple ways to accomplish the same output data structure, and a certain degree of regularity is needed to simplify the data conversion. Once these rules are set in place, creating the schemas is relatively straightforward. Figure 5 shows an example of XML schema syntax.

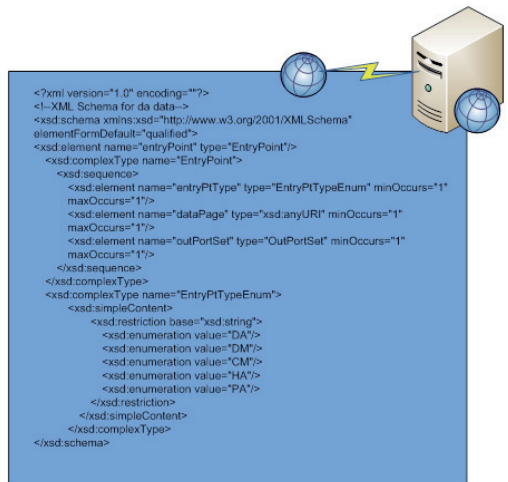


Figure 5 XML schema of transformed and transferred data

All existing data (assets, events, failures, alarms) can be modelled with XML. Among them, the most critical and difficult to represent is the layer that represents information on sensory inputs and outputs, whether a single scalar value or an array of complex data points. The standards define various data formats that may be implemented to represent sensory information. Sensory data, especially relevant in condition monitoring and process control, may be as simple as a single value or as complex as several synchronous sampled waveforms.

With this system, each data originator can wrap its data using a schema understood by or convenient for that device or application, and each receiving application can receive the data in a different schema used for or understood by the receiving application. The XML server may be configured to map one schema to another schema depending on the source and destination(s) of the data. It may also perform certain data processing functions or other functions based on the receipt of data. The mapping and processing function rules are set up and stored in the server prior to the operation of data integration applications. In this manner, data may be sent from any one application to one or more other applications.

3.3 OPTIRAIL: Asset Data Integration Using XML

Web based technologies have been widely used for eMaintenance purposes according to [8], [14], [15]. One architecture for collecting and integrating data from disparate data sources using an XML server based on web services is proposed in Figure 6.

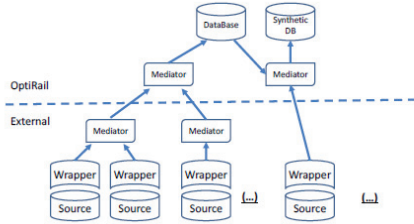


Figure 6. Integration of disparate data sources

In this model, it is understood the data collected from disparate sources are converted into a common format using XML. In order to enable data from different data sources to be collected and used in a single system as a single data source, a configuration database or other integrated configuration system must be provided. An explorer type display or hierarchy should be provided to allow the manipulation, organisation and use of the collected data to make those data available to different applications.

Figure 6 illustrates an architectural overview of a system which implements the collection of data from disparate data sources with a process control system. The system may include a maintenance management system, a product inventory control system, a production scheduling system, and other systems connected by a LAN, the Internet, etc. XML is used as the transaction server. The

server sends XML wrapped data to the web services indicative of the data.

The web services must include a series of web service listeners, which listen for or subscribe to certain data from other data sources and provide these data to the subscribing applications. The web listening services (which may be part of the data collection and distribution system) may listen for and redistribute alarms and events data, process condition monitoring data and equipment condition monitoring data. Interfaces for these data are used to convert the data to a standard format or protocol, such as the Fieldbus or XML, as desired.

The web services will be in contact with and receive data from other external data sources via web servers. These external sources may include vibration monitoring data sources, real-time optimisation data sources, expert system analysis data sources, predictive maintenance data sources, loop monitoring data sources, or other data sources.

Finally, a configuration database is used to store and organise the data from the process control runtime system, including any data from the remote data sources, such as external web servers.

3.4 System Context

The OPTIRAIL system framework supports the integration of various data sources as expected for context creation and exploitation. These data sources could take different formats and natures, as noted in the different case studies. To handle these differences, the OPTIRAIL framework should provide facilities for data wrapping and mediation at the OPTIRAIL level, along with interfaces for external data wrappers and mediators

As a framework, OPTIRAIL should allow the addition of new sources and mediation procedures, along with the necessary data validation and consistency checking needs. From the OPTIRAIL operation point of view, different data spaces are managed at different levels of the system.

At the data management data space, the following agents and data bases are managed and merged for context production:

- Database, containing the database baseline;
- Synthetic database, containing derived calculations from the database or from external sources not included in the database;
- Database management information;
- Wrapper and mediator management information;
- Archived data.

4. DETECTING CONTEXT ANOMALIES

4.1 Contextual Failures

If a data instance is anomalous in a specific context, but not otherwise, it is termed a contextual anomaly (also called a conditional anomaly [11]). The notion of a context is induced by the structure in the dataset and has to be specified as a part of the problem formulation. Each data instance is defined using the following two sets of attributes: contextual and behavioural. Contextual attributes are used to determine the context (or neighbourhood) for that instance. For example, in spatial datasets, the longitude and latitude of a location are the contextual

attributes. In time-series data, time is a contextual attribute that determines the position of an instance on the entire sequence. Behavioural attributes define the non-contextual characteristics of an instance. For example, in a spatial data set describing the average rainfall of the entire world, the amount of rainfall at any location is a behavioural attribute. The anomalous behaviour is determined using the values for the behavioural attributes within a specific context. A data instance might be a contextual anomaly in a given context, but an identical data instance (in terms of behavioural attributes) could be considered normal in a different context. This property is key in identifying contextual and behavioural attributes for a contextual anomaly detection technique. Contextual anomalies have been explored in time-series data [16; 9] and spatial data [5; 10].

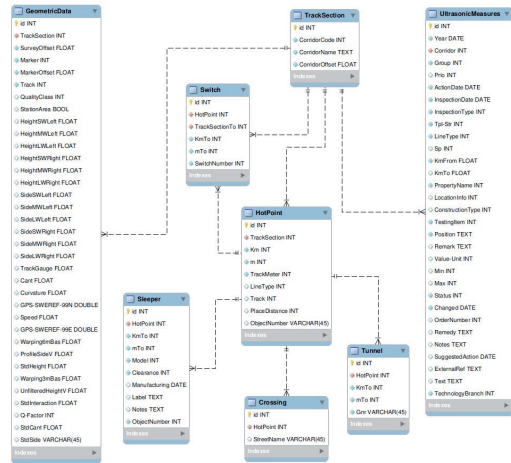


Figure 7. Database design for Sweden case study

Figure 7 shows a temporal work order histogram for Sweden. From the figure, we see that work is often performed on various parts of the Iron Ore Line during the same months of the year. In certain months (likely in the winter), no work is done. Apparently, no work was done on part 111 after 2009; it is unclear, however, whether no work was done by the contractor or if the contractor simply did not notify TRV of work done.

Combining contextual information from weather databases and work orders allows users to normalize the WOs by taking into consideration climate conditions and identifying situations as normal in the Swedish context; i.e., no work is scheduled during winter. In this scenario, the lack of work orders is not considered anomalous. In southern countries, this would definitely be abnormal behaviour, as the distribution of work is entirely different. In summary, a contextual approach can help users detect both anomalies and false alarms, i.e., situations that are normal in the current context but anomalous elsewhere.

4.2 Using Fuzzy Logic to Solve Diagnosis Problems

A system able to detect, isolate and identify faults is called a fault diagnosis and isolation system (FDI). Over the years, a great deal

of research has used analytical approaches based on quantitative models to generate signals that reflect inconsistencies between normal and faulty system operation. The detection of context inconsistencies comprised of several signals and texts requires an approach able to handle information reduction and fusion so that information entities can be identified and tracked down. Early detection and isolation of abrupt and incipient contextual faults can be achieved with model-based processing of all measured variables, using either qualitative or quantitative modelling.

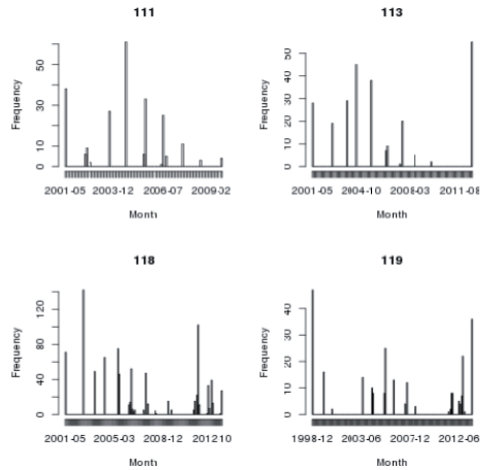


Figure 8. Temporal work order histogram for Sweden

Generally speaking, there are two types of fault detection: process variable monitoring and model based methods, the latter of which are more complex. For a simple fault that can be detected by a single measurement, a conventional threshold check may be appropriate. However, since in complex systems it is usually very difficult to directly measure the state of the process, more sophisticated solutions are needed. In this case, a model-based approach is more suitable, but this requires process modelling, a very demanding task.

The idea of model-based fault detection is to compare output signals of the model with the real measurements available in the process, thereby generating residuals or fault indicators giving information about the location and timing of a fault. The approach requires precise mathematical relationships relating the model to the process if small abrupt and incipient faults are to be detected quickly and reliably.

Fuzzy techniques have received much attention because of their fast and robust implementation, their capacity to embed a priori knowledge, their performance in reproducing nonlinear mappings, and their ability to generalise. Fuzzy logic techniques are now being investigated in the FDI research community as a powerful modelling and decision-making tool, along with neural networks and other more traditional techniques, such as non-linear and robust observers, parity space methods and hypothesis-testing theory. To circumvent the problem of precision modelling, more abstract models based on qualitative approaches may be used;

these are a better fit for contextual thinking since exact equations and accurate physical models are not an option in extremely complex systems.

Alternatively, fuzzy-logic rules may be developed to either assist or replace the use of a model for diagnosis. The main advantage of fuzzy logic is that it enables the system behaviour to be described by “if-then” relations, ideal for contextual segregation and selection. The main trend in developing fuzzy FDI systems has been to generate residuals using either parameter estimations or observers and allocate the decision-making to a fuzzy-logic inference engine. By doing so, it has been possible to combine symbolic knowledge with quantitative information, thereby minimising the false alarm rate.

A key benefit of fuzzy logic is that it lets the operator describe the system behaviour or the fault-symptom relationship with simple if-then rules. In this paper, we take another step forward. More specifically, we generate symptoms using fuzzy observers and measurements. The underlying idea is to predict the system outputs from the available inputs and outputs of the process, thus identifying a fuzzy model directly from data. The residual will then be a weighted difference between the predicted and the actual outputs. In our approach, fuzzy observers are built for normal and faulty operations, allowing the detection and isolation of the considered faults [4].

4.3 The Need For Complex Relations In Contextual Decision Making.

As the key objective of applied fuzzy systems is to transfer ambiguity into value, the main application effort is the translation of the vague information from the natural language of the experts into the precise and highly interpretable language of fuzzy sets and fuzzy rules. Experts play a leading role in this process even in the case of data-driven fuzzy systems; although the rules are automatically discovered by the clustering algorithms, they must be interpreted by the experts [7].

Maintenance engineering and management face a number of problems in diagnosis and prognosis when decisions must be made, and certain issues related to the quality and quantity of information justify fuzzy and contextual approaches:

- Vague knowledge must be included in the solution;
- The solution must be interpretable in some form of linguistic rules, i.e. we want to learn about our data/problem;
- The solution must be easy to implement, use, and understand;
- Interpretation is as important as performance.

Information quality and quantity are important to the decision making process when a fuzzy system is used. Even though there is no fixed order to follow when designing a fuzzy system, an attempt to define an application sequence for classical expert-based systems is given in Figure 8.

Probably 80% of the application success depends on the efficiency and quality of knowledge acquisition. This is the process of extraction of useful information from the experts, datasets, known documents and common sense reasoning applied to a specific objective. It includes interviewing the experts and

defining key features of the fuzzy system, such as: identifying input and output variables, separate crisp and fuzzy variables, formulating the proto-rules using the defined variables, ranking the rules according to their importance, identifying operational constraints, and defining expected performance.

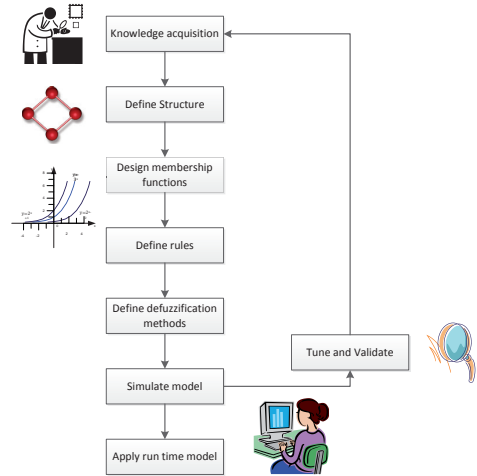


Figure 9. Application sequence for classical expert-based systems (Kordon, 2010)

Just as the success of expert-based fuzzy systems depends on the quality of knowledge acquisition, so too the success of data-based fuzzy systems is linked to the quality of the available data.

Data collection is the most critical part of the process; we will have a problem if the data have very narrow ranges and the process behaviour cannot be represented adequately. The chance for appropriate fuzzy rule discovery is very low in this case.

Once data are collected, in the next step, data processing, we look for proto-clusters and try to define the corresponding rules. The most interesting step is determining the size of the granule of the proto-clusters. In principle, the broader the cluster space, the more generic the defined rule. However, some important nonlinear behavior of the process could be lost. It is recommended that the proper size of the fuzzy clusters be decided by domain experts.

The result of the development process is a fuzzy system model based on the generalised rules. There is no difference in the runtime application between either approach (Kordon, 2010).

5. CONCLUSION

To create a more efficient railway, we suggest using context engines to define the existing interactions within the railway system. Because there are many different maintenance data sources, i.e. different nature, origin, granularity, decision making can be problematic. The only feasible solution seems to be a contextualisation of the data in such a way that condition monitoring data, i.e. physical measurements and other variables, can be understood under certain conditions (weather, regulations

etc.) and as a consequence of certain actions (maintenance interventions, overhauls, outsourcing warranties) .

If railway maintenance is context driven, users will better understand the meaning of physical information in certain scenarios where the surrounding environment or collateral events could have an effect. For example, identifying false alarms and diagnosing outlier behaviour are two of the main advantages of the contextual approach.

Most of the railway's context information comes from work orders and experience, making it qualitative rather than quantitative. The fuzzy logic approach can integrate such information sources for diagnosis and proves to be a fast and robust way to deal with these fluffy data. Transferring ambiguities typical in maintenance into numerical values will improve the accuracy and quality of decision making in the maintenance of the railway infrastructure.

6. ACKNOWLEDGMENTS

The research has received funding from the European Community's Framework Programme FP7/2012-2015 under grant agreement no. "314031". The Consortium consists of VIAS, SINTEF, EVOLEO, CARTIF, UGR, OSTFALIA, MERMEC, ADIF, and LTU.

7. REFERENCES

- [1] Camci, F., Valentine, G.S., Navarra, K., (2007). "Methodologies for integration of PHM systems with maintenance data." *In: Aerospace Conference*. IEEE
- [2] H. Min-Hsiung, C. Kuan-Yii, H. Rui-Wen, C. Fan-Tien, (2003) "Development of an ediagnostics/ maintenance framework for semiconductor factories with security considerations", *Advanced Engineering Informatics* 17 (3-4) pp.165-178.
- [3] I. Hausladen, C. Bechheim. (2004) "E-maintenance platform as a basis for business process integration", in: 2004 2nd IEEE *International Conference on Industrial Informatics*, 2004 (INDIN'04), June 24-26, pp. 46-51.
- [4] Kordon, A. K., (2010). "Applying Computational Intelligence: How to Create Value". Berlin, Heidelberg: Springer-Verlag.
- [5] Kou, Y., Lu, C. & Chen, D. (2006). "Spatial Weighted Outlier Detection". Bethesda, Maryland, u.n., pp. 613-617.
- [6] M. Wollschlaeger, T. Bangermann, (2003) "A Web-based maintenance portal based on an XML content model", in: *IEEE International Conference on Industrial Technology*, vol. 1, 2003, pp. 405-410.
- [7] Mendonça, L. F., Sá da Costa, J. M. G. & Sousa, J. M. (2003). "Fault detection and diagnosis using fuzzy models". u.o., u.n., pp. 1-6.
- [8] Parida, A. and Kumar, U. (2004). "Managing Information is the key to Maintenance Effectiveness", *e-Proceedings of Intelligent Maintenance System*, Arles, France, pp.15-17
- [9] Salvador, S. & Chan, P. (2003). "Learning states and rules for time-series anomaly detection", Melbourne: u.n.
- [10] Shekhar, S., Lu, C. & Zhajng, P.,(2001). "Detecting graph-based spatial outliers: Algorithms and applications (a summary of results)". i: *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. u.o.:ACM Press, pp. 371-376.
- [11] Song, X., Wu, M., Jermaine, C. & Ranka, S. (2007). "Conditional anomaly detection". *IEEE Transactions of Knowledge data Engineering*, 19(5), pp. 631-645.
- [12] Szymanski, J., Bangemann T., Thron M., Thomesse, J.P. , Reboeuf, X., Lang C., Garcia, E. (2003), *PROTEUS*-"a European initiative for e-maintenance platform development", *Conference on Emerging Technologies and Factory Automation*. Proceedings, vol. 2, pp. 415-420.
- [13] T. Han, B.S. Yang, "Development of an e-maintenance system integrating advanced techniques" (2006), *Computers in Industry* 57 (6) pp.569-580.
- [14] U Kunze, "Condition tele monitoring and diagnosis of power plants using Web technology" (2003), *Progress in Nuclear Energy* 43 pp.129-136.
- [15] V.M. Catterson, E.M. Davidson, S.D.J. McArthur. (2005) "Issues in integrating existing multi- agent systems for power engineering applications", in: *Proceedings of the 13th International Conference on Intelligent Systems Application to Power Systems (ISAP'05)*, pp. 6.
- [16] Weigend, A., Mangeas, M. & Srivastava, A. (1995). "Nonlinear gated experts for time-series: Discovering regimes and avoiding overfitting". *International Journal of Neural Systems*, 6(4), pp. 373-399.

eMaintenance 1B

A Coloured Petri Net Model of Fleet Cannibalisation

Jingyu Sheng
University of Nottingham
University Park, Nottingham, UK
evxs6@nottingham.ac.uk

Darren Prescott
University of Nottingham
University Park, Nottingham, UK
darren.prescott@nottingham.ac.uk

ABSTRACT

Cannibalisation is a maintenance activity that involves removing serviceable parts from inoperative platforms to replace unserviceable parts of the same type in other platforms. It can provide a significant benefit to fleet readiness, particularly if spare parts are in short supply. However, cannibalisation also has drawbacks: it brings an increased workload for maintenance crews and parts can be damaged during the cannibalisation process. For this reason, it is important to have a clear understanding of the effects that cannibalisation will have on fleet operation and maintenance. Accurate models are needed to predict the effects of cannibalisation on fleet performance and to provide fleet managers with trustworthy information on which to base maintenance decisions relating to cannibalisation and spare parts provision. This paper presents a coloured Petri net (CPN) model of fleet cannibalisation that takes account of fleet operation and a number of factors relating to maintenance. An example fleet is modelled and measures of average fleet readiness and maintenance cost are used to evaluate the effects of cannibalisation on fleet performance.

Keywords

Fleet, Maintenance, Cannibalisation, Coloured Petri Net (CPN)

1. INTRODUCTION

Cannibalisation involves removing serviceable parts from a platform which has failed and using them to repair other unserviceable platforms in the fleet. It can be used to maintain the operational capability of a fleet of similar platforms when replacement parts are difficult to obtain or their provision is delayed. Cannibalisation is seen as a necessity in many organisations since it can significantly reduce the maintenance turnaround for a failed platform when there is a lack of spare parts. For example, a recent study documented over 850000 cannibalisation actions performed over a five-year period in the U.S. Air Force and U.S. Navy, with these actions taking 5.5 million maintenance man-hours [3]. This high workload is one of the main drawbacks of cannibalisation, along with the potential for damage to occur on cannibalised units during the cannibalisation process. Similar fleet readiness benefits to those obtained by applying cannibalisation can be gained by increasing spare part inventories; however, this can lead to significant cost increases. Therefore, the trade-off between the use of cannibalisation and spare part inventories is a complex problem for maintenance decision makers to address.

Current research on cannibalisation mainly focuses on evaluating the impact of cannibalisation on fleet performance and spare part inventories. Fisher [1] provides an overall survey of the issues present in maintenance decisions involving cannibalisation along with a review of relevant models. Fisher [2] develops an analytical model based on Markov processes to study and

compare the effects of different cannibalisation policies. Manpower constraints are considered, represented by the assumption that there is only a single repair facility in the fleet. The studied platforms are identical and consist of just two components connected in series, with each component having a constant failure rate in order to satisfy the Markov assumptions.

Cassady et al. [3] develop a discrete-event simulation model to evaluate the impact of cannibalisation on the readiness and average manpower working hours of a fleet. In common with the work presented by Fisher [2], each platform in the fleet is assumed to be comprised of two components connected in series. Ormon and Cassady [4] extend this work by taking a discrete-event simulation model as a decision-support tool to optimise cannibalisation policies and spare part inventories. Two different optimisation models are established. The first minimises maintenance cost subject to a fleet readiness constraint and the second maximises fleet readiness subject to maintenance investment constraints. The platforms studied are the same as those studied by Cassady et al. [3].

Salman et al. [5] extend the work of Ormon and Cassady [4] in two ways. Firstly, two different methods of selecting both the platform to be cannibalised and the platform to be restored by cannibalisation are studied and compared. Secondly, a cannibalisation-induced damage model is established to evaluate the reduced residual life of functioning components caused by cannibalisation. Further, the simulation model is applied to a fleet of identical platforms, each of which contains five components connected in series.

This paper presents a novel coloured Petri net (CPN) model of the fleet cannibalisation process and a number of factors relating to fleet maintenance: spare part inventory, repair, platform failure logic and queuing of platforms requiring maintenance. The CPN model consists of four modules: a fleet module, a platform module, a component module and a maintenance module. The model is constructed in such a way that complex platform architectures are simple to model and, since it is modular, it can be easily extended in future to consider other aspects of cannibalisation and maintenance in order to provide a practical decision-making tool for fleet operators and maintenance engineers. As well as providing clearly-auditable, concise models, CPN form an ideal framework for Monte Carlo simulation. In this paper, the use of the CPN model to study the effect of cannibalisation on fleet performance is demonstrated for a simple fleet.

2. PETRI NETS

First introduced by Carl Adam Petri [7], Petri nets (PN) are powerful graphical and mathematical tools for modelling complex, dynamic systems. A PN is a directed graph with two types of nodes: *places*, shown as circles; and *transitions*, drawn as bars. Different types of nodes are connected to one another by directed

edges or arcs. Places contain a discrete number of tokens, represented graphically by bullet points. The distribution of tokens within the PN defines its *marking*, which describes the state of the modelled platform at any point in time. Transitions govern the dynamic behaviour of the PN; the platform state changes when the process of *switching* or *firing* of any transition happens. In order to fire, a transition must first be *enabled*, which occurs when the number of tokens in each input place is no less than the *weight* of the arc that links it to the transition. Transitions can be *immediate*, represented by solid bars, or *timed*, represented by hollow bars. Immediate transitions fire as soon as they are enabled and timed transitions fire after a *delay* has elapsed following their becoming enabled.

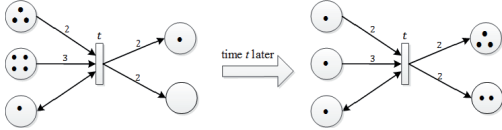


Figure 1. Transition enabling and firing

An example of transition enabling and firing is shown in Figure 1. The transition shown in Figure 1 is a timed transition with three input places and three output places (the place linked by the double-headed arc is both an input and output). A double-headed arc is shorthand for two arcs of equal weight (shown next to each arc if not equal to 1) leading in each direction between a place and a transition. The transition is enabled since the number of tokens in each input place is no less than the weight of the arc that links it to the transition. After a delay time t has elapsed a number of tokens equivalent to the associated arc weight is removed from each input place and added to each output place. The delay time t can be determinate or can be sampled from a distribution.

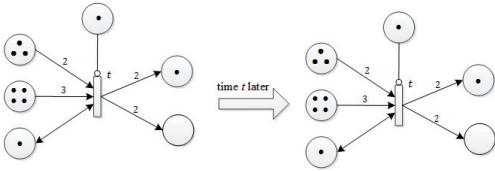


Figure 2. Inhibitor arc prevents transition firing

If a transition is linked to an input place by an inhibitor arc, its firing process is prohibited when the number of tokens in the place is greater than or equal to the weight of the inhibitor arc. An inhibitor arc is drawn with a circle at its head instead of an arrow. Figure 2 shows a situation where a transition that would otherwise be enabled, is not enabled due to the marking of a place linked to the transition by an inhibitor arc.

Jensen and Kristensen [6] created coloured Petri nets (CPN) as an extension of PN. CPN allow more concise system models to be built than would be possible by using PN. In CPN, each token is assigned a label, called the *token colour*. The colour is a graphical method used to distinguish between different tokens. The token colour can represent different data types, such as Boolean, integer or character strings. Each place is assigned a *colour set*, which is the set of possible token colours of a specified type. Transitions fire in different ways depending on tokens' colours. The colour of a token can change when it passes through a transition, as decided by a specific modelling condition.

The enabling and firing policies of a transition in CPN are governed by *arc expressions* and transition *guards*. Arc expressions, shown as textual inscriptions next to arcs, determine the amount and colours of tokens that are removed from input places and added to output places when a transition fires. The guard of a transition is a *Boolean* expression that represents a constraint on its enabling policy. A transition is only enabled when the marking of its input places satisfies the input arc expressions and the guard of the transition evaluates to be true. Figure 3 shows an instance of transition enabling and firing for a CPN.

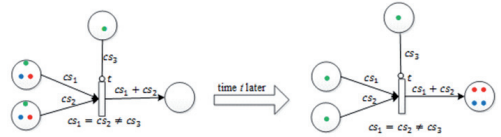


Figure 3. Transition enabling and firing of CPN

Figure 3 shows a timed transition with three input places and one output place. All four places are assigned with the same colour set, $CS = \{\text{green, red, blue}\}$. Variables cs_1 , cs_2 and cs_3 are colours that belong to CS . The transition is enabled when there are tokens in its three input places that satisfy its guard $cs_1 = cs_2 \neq cs_3$. After a delay time t , the transition switches red and blue tokens from the input places to the output place. The switching of green tokens is not enabled since the place linked by the inhibitor arc contains a green token.

3. CPN CANNIBALISATION MODEL

This section outlines the main assumptions that are made with regard to how cannibalisation is applied to a fleet. An overview of the structure of the CPN cannibalisation model is then presented.

3.1 Fleet Definition

A fleet consists of K identical and independent platforms, each of which contains N components. The time to failure for each component follows a known distribution and each component (and hence platform) can only fail during operation. Each platform can be assigned to perform missions only when all of its components function. i and j represent a particular platform in the fleet and a particular type of component respectively. Component j of platform i is represented by (i, j, a) where $a (a \geq 0)$ is the age (in working hours) of component type j of platform i . A spare component of type j is represented by $(0, j, 0)$. If a platform fails because of component failures, it will be immediately taken out of service and added to the 'platform maintenance queue' which holds failed platforms that are required to be maintained and restored to the working state. If there is more than one failed platform in the platform maintenance queue, failed platforms are maintained according to an order determined by, for example, random selection (RS), first-in-first-out (FIFO) or last-in-first-out (LIFO). Maintenance involves technicians firstly removing all failed components and then using working components to restore the platform to the working state.

When a platform is in the platform maintenance queue, it is first checked to see whether it can be entirely restored to the working state using components from the spare part inventory. If so, maintenance technicians will install the required spare components in order to return the platform to the working state. Otherwise, if cannibalisation is permitted, technicians will check

whether there are available working components in other failed platforms. If there are, cannibalisation is performed by removing the required serviceable components from other failed platforms and installing them in the platform being maintained. A combination of spare parts and cannibalised components may be needed to repair a failed platform. When the repair is finished, the platform can again be assigned to perform missions. If the availability of spare and cannibalised components is such that the failed platform cannot be entirely restored to the working state, maintenance of the platform will not begin and the platform will be added to the 'platform waiting maintenance queue' which holds all failed platforms for which there are insufficient resources to complete the platform repair.

Once removed, failed components are repaired at repair shops before being returned to the fleet. These components form the new spare part inventory. If there are failed platforms in the 'platform waiting maintenance queue' when new spares arrive, technicians check the queue to see if any of the failed platforms within it can be entirely restored using these new spares and those already in the spare part inventory. If so, maintenance crews use the spares to perform the required maintenance. Cannibalisation is also performed if necessary. If the waiting platforms cannot be entirely restored using the available spares and cannibalised parts then those platforms will remain in the 'platform waiting maintenance queue' until the required spares become available. Manpower limitations are not considered in this paper, meaning that maintenance personnel are always available to perform the required maintenance.

3.2 CPN Model Structure

The CPN cannibalisation model is made up of four modules:

- *Fleet module* – models the fleet operation;
- *Platform module* – models platform failures;
- *Component module* – models component failures;
- *Maintenance module* – models cannibalisation and other maintenance factors including spare inventory, repair of failed components and platform restoration queuing.

Figure 4 shows the relationship between the four modules. The component module outputs component failures to the platform module. If component failures lead to a platform failure, the platform module, in turn, links the platform failure to the fleet module, which takes the failed platforms out of service and outputs them to the maintenance module. After all the failed components of a failed platform are replaced by spare parts or through cannibalisation, the restoration of the platform is finished. The maintenance module ensures the replacement of all failed components in a failed platform by spares or cannibalised components and outputs working platforms and components to the fleet module and component module respectively. The separate modules are described below. The colour sets and colour variables applied in these modules are defined in Tables 1 and 2 respectively. For an arbitrary variable c of the colour set C , $c(i)$, $c(j)$ and $c(a)$ represent the platform index, component type and component age respectively.

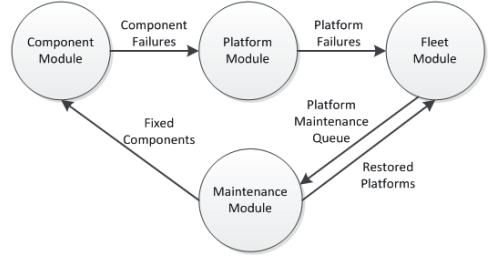


Figure 4. Interactions between modules

3.3 Fleet Module

The fleet module models the operational state of all platforms in the fleet. If a working platform fails, it is immediately taken out of service and added to the platform maintenance queue, from where it will be maintained if the parts required to restore the platform to the working state are available. Figure 5 shows the fleet module. Places P_{PW} and P_{PF} represent working and failed platforms respectively. Places P_{PPM} and P_{PMQ} respectively represent platforms that are performing missions and in the platform maintenance queue.

It is assumed that all working platforms are assigned to perform missions. Therefore, if P_{PW} is marked the transition T_{PPM} is enabled and fires immediately, moving all tokens in P_{PW} to P_{PPM} .

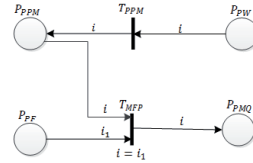


Figure 5. Fleet Module

As soon as a platform fails it is taken out of service and added to the platform maintenance queue. This is modelled by transition T_{MFP} . If a platform i fails while in service (both P_{PF} and P_{PPM} have a token with colour i), T_{MFP} is enabled and will fire immediately, removing tokens with colour i from both P_{PF} and P_{PPM} , and creating a token with the same colour in P_{PMQ} .

3.4 Platform Module

The platform module models the process of platform failure. The different platform failure modes are modelled using fault trees, which show how platform failure, represented by a top event, is caused by component failures, represented by basic events. The fault tree is constructed by breaking down the top event into its immediate, necessary and sufficient causes, using logic gates and intermediate events. The causes of the intermediate events are then determined in the same way. This process continues until the basic events are reached. Figure 6 shows an example fault tree for a platform failure mode, represented by the event TOP . There are ten basic events, numbered 1-10, four intermediate events, labelled IE_1 - IE_4 and five logic gates, labelled G_1 - G_5 (G_1 and G_2 are examples of OR and AND gates respectively).

Table 1. Colour sets

Colour set	Colours	Meaning
I	$\{1, 2, \dots, K\}$	Platform index
J	$\{1, 2, \dots, N\}$	Component type
A	$a \geq 0$	Component age: the number of cumulative working hours
C	$\{(i, j, a)\}$	Components: $i=0$ means spare component.
CRD	$\{(c, c_1)\}$ $\forall c, c_1 \in C, c(j)=c_1(j), c(i) \neq 0$ and $c(i) \neq c_1(i)$	Component replacement decision: the inoperative component c is decided to be replaced by component c_1 . $c_1(i)=0$ means c will be replaced by a new spare component. $c_1(i) \neq 0$ means that the working component c_1 is cannibalised from platform $c_1(i)$ to replace the failed component c in platform $c(i)$.

Table 2. Colour variables

Variable	Colour	Meaning
i, i_1	I	A platform
j	J	A component type
c, c_1, \dots, c_n	C	A component
crd	CRD	A component replacement decision
a	A	The age of a component

Figure 7 shows the CPN that is equivalent to the example fault tree shown in Figure 6. Each non-basic event in the fault tree is represented by a place with colour set I . The basic events, which would usually represent component failures, are represented by the place P_{CF} , which has colour set C . AND gates are converted to immediate transitions, with a single input place, P_{CF} , since all basic events are represented by this place. OR gates are represented by a number of immediate transitions equal to the number of places used to represent the gates' input events.

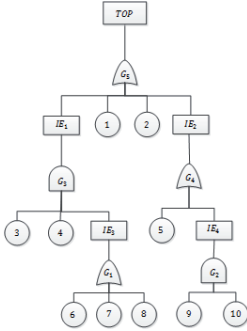


Figure 6. Example platform fault tree

If component failures lead to the failure of platform i then tokens with colour i will exist in P_{TOP} and in P_{PPM} , enabling the transition T_{PF} . T_{PF} will then fire immediately and create a token of colour i in P_{PF} , representing the fact that the platform is failed and that maintenance is required immediately. The transitions linking P_{PW} and the places that represent non-basic events are designed to clear all tokens representing the fault tree intermediate events of a failed platform from the platform module after the platform is restored to the working state.

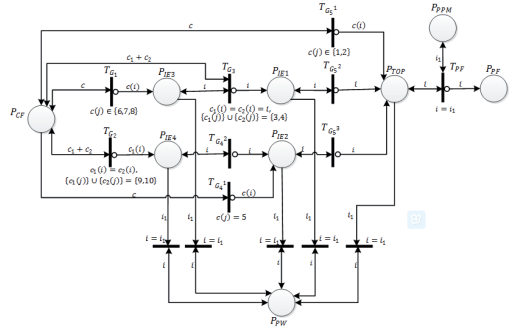


Figure 7. Example platform module

3.5 Component Module

Component failures are modelled using the component module shown in Figure 8. It is assumed that components can only fail when the platforms in which they are fitted are performing missions and that the time to failure of each type of component follows a known distribution.

P_{CW} represents the component being in the working state and P_{CF} represents the component being in the failed state. Transition T_{CF} is enabled when there is a token with colour c in P_{CW} and one with colour i in P_{PPM} which satisfies $i = c(i)$, meaning that the platform in which the component is fitted is performing a mission. If T_{CF} is enabled, a firing delay representing the time to component failure is obtained by randomly sampling the appropriate failure time distribution given the component type $c(j)$ and age $c(a)$.

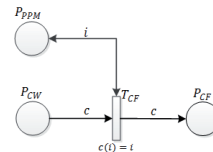


Figure 8. Component module

Each component ages while in operation. This is accounted for in the model by increasing the age, $c(a)$, of the token with colour c while it enables T_{CF} . If T_{CF} is still enabled by the token when the sampled delay time has elapsed, T_{CF} switches the token with

colour c from P_{CW} to P_{CF} which means that component c fails. If the platform $c(i)$ is taken out of service because of other component failures (meaning that the token with colour $c(i)$ is removed from P_{PPM}) before the delay time runs out, the firing process of the token by T_{CF} is interrupted, ensuring that inoperative components cannot experience failures. As T_{CF} is disabled, the component age, $c(a)$, stops increasing.

3.6 Maintenance Module

The maintenance module models the selection of failed platforms to be maintained and the selection of components that are used to return failed platforms to the working state, either from the spare part inventory or from other failed platforms, as well as the component maintenance actions themselves. This module is itself divided into four sub-modules:

- Platform Restoration Queuing (PRQ) sub-module,
- Component Replacement Decision (CRD) sub-module,
- Component Maintenance Actions (CMA) sub-module,
- Cannibalisation Inventory Control (CAIC) sub-module.

3.6.1 Platform Restoration Queuing

The PRQ sub-module selects failed platforms from the maintenance queue to be considered for maintenance, according to some queuing discipline specified by the maintenance manager, such as first in first out (FIFO), last in first out (LIFO), random selection (RS), etc.

Figure 9 shows the PRQ sub-module. P_{PMQ} is marked with tokens representing the failed platforms in the maintenance queue. Transition T_{SP} expresses the maintenance sequence of inoperative platforms and is enabled when P_{PMQ} is marked. Once enabled it fires immediately, marking P_{SPR} with a token of colour i ; this represents the fact that platform i has been selected to be restored to the working condition. However, the inhibitor arc from P_{SPR} to T_{SP} ensures that there is at most one token in P_{SPR} at a time, ensuring that the spare part and cannibalisation inventories are built separately for each platform.

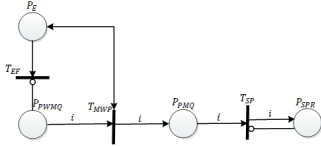


Figure 9. PRQ sub-module

Maintenance of a failed platform can only begin if all of its failed components can be replaced by either spares or cannibalised components; this condition is checked in the Component-Replacement Decision (CRD) sub-module. If it is not possible to fully restore the platform to the working state then maintenance of the platform must be delayed until the appropriate parts become available, either as spare parts or through cannibalisation. In this case the CRD sub-module ensures that the token is removed from P_{SPR} and one of the same colour is added to place P_{PMMQ} , which represents failed platforms that are waiting for parts to become available before maintenance can be performed, the platform waiting maintenance queue. If P_{PMQ} is still marked after the token in P_{SPR} is removed then T_{SP} is enabled again and fires immediately, meaning that another failed platform will be checked to see whether it can be fully restored to the working state.

If there are platforms that are waiting for spare parts to become available (P_{PMMQ} is marked) when new spares become available (P_E is marked), T_{MWP} is enabled and fires immediately. All tokens in P_{PMMQ} are immediately moved to P_{PMQ} because of the double-headed arc between P_E and T_{MWP} , before T_{EF} absorbs the token in P_E . This ensures that when new spare parts become available all platforms are checked to see if they can be fully restored to the working condition.

3.6.2 Component Replacement Decision

The CRD sub-module, shown in Figure 10, determines whether failed platforms can be fully restored to a working condition through the use of spares or cannibalisation and selects the parts that will be used to replace failed components during maintenance. If parts are available to replace all failed components the platform can be restored to the working condition; if not, the platform is placed in the queue for platforms that are waiting for maintenance. Any failed platforms that are being used as a source of cannibalised components are placed directly in this queue if the removal of cannibalised components from them is incomplete. This ensures that a platform cannot be selected for maintenance if it is being cannibalised. The meanings of transitions in the CRD sub-module are listed in Table 3.

If spare parts are available to replace failed components, maintenance crews choose these in preference to cannibalising components from other failed platforms. If a number of spares are available then a queuing discipline is applied to select the one to be used during maintenance.

When platform i is selected for possible restoration by the PRQ sub-module (which involves marking P_{SPR} with a token of colour i), it must be determined whether or not maintenance can be performed to fully restore platform i to the working state. If platform i is being used as a cannibalisation source then there will be at least one token with colour c and $c(i) = i$ in place P_{CCAQ} . This will enable T_{CAPW} , which will remove the token from P_{SPR} and place an identical token in P_{PMMQ} (meaning that platform i cannot be maintained and is placed in the queue for platforms that are waiting for maintenance).

If platform i is not a cannibalisation source then it must be determined whether or not parts are available to replace all failed components in the platform. The failed components are represented as tokens in P_{CRPQ} . Transition T_{SFC} is enabled and fires immediately, marking P_{CSPR} , which has colour set C , with tokens representing the failed components of platform i .

Now that the failed components of platform i have been identified, by marking P_{CSPR} , the availability of the required spare and cannibalised parts to complete the restoration of the platform to the working state must be checked. Two lists are produced. The first is a list of spare parts that are available to replace failed components and the second is a list containing those parts that can be cannibalised from other failed platforms but for which spares are not available. If, for any component, a replacement cannot be added to either list then complete restoration of the platform to the working state is impossible and maintenance will not begin.

Transition T_{PSID} is responsible for compiling the list of components that can be replaced by parts from the spare part inventory. T_{PSID} is enabled when a component in the spare part inventory (represented by a token in P_{SCI}) is of the same type as a failed component in platform i (represented by a token in P_{CSPR}).

When it fires, a token is removed from both the list of failed components and the spare part inventory and added to P_{PSID} , meaning that the decision has been made to replace this failed component by a spare part if maintenance of this entire platform is possible.

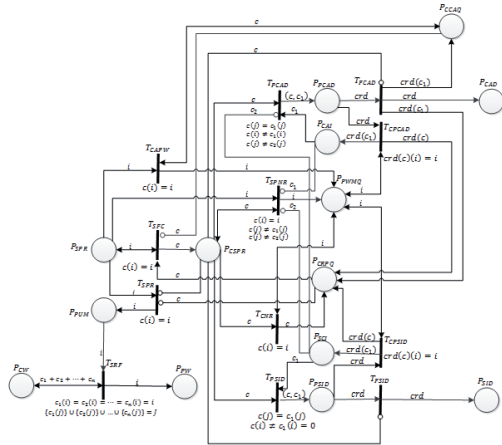


Figure 10. CRD sub-module

Table 3. Meanings of transitions in the CRD sub-module

Transition	Meaning
T_{SFC}	Identify all failed components of the selected failed platform that must be restored
T_{CAPW}	Add the cannibalisation source platforms to the platform waiting maintenance queue
T_{SPR}	Identify the selected platform to be restorable
T_{PCAD}	Make primary component cannibalisation decisions
T_{SPNR}	Identify the selected platform to not be restorable
T_{CNR}	Failed components cannot be replaced due to lack of resources
T_{PSID}	Make primary spare installation decisions
T_{FCAD}	Set final component cannibalisation decisions
T_{CPCAD}	Clear primary component cannibalisation decisions
T_{CPSID}	Clear primary spare installation decisions
T_{FSID}	Set final spare installation decisions
T_{SRF}	Inoperative platforms restored

T_{PCAD} is responsible for compiling the list of components for which spare parts are unavailable but which can be replaced by cannibalised components that are available in the cannibalisation inventory. If a part is available in the cannibalisation inventory (token in P_{CAI}), which is of the same type as a failed component in the platform (represented by a token in P_{CSPR}) then T_{PCAD} is enabled and fires, leading to a token representing the type of failed component being removed from both P_{CSPR} and P_{CAI} and an equivalent token added to P_{PCAD} , meaning that the decision has been made to replace this failed component by a cannibalised part if maintenance of this entire platform is possible.

If no parts, either spare or cannibalised, are available to replace any of the failed components in platform i then maintenance of

the entire platform is not possible and will not go ahead for this platform. In this case the platform is placed in the queue of platforms that must wait for maintenance to take place. This process is governed by T_{SPNR} , which removes the token representing the platform from P_{SPR} and places an identical token in P_{PVMQ} . This enables T_{CNR} , which takes all remaining tokens with colour c from P_{CSPR} and puts them in P_{CCRPQ} , indicating that all failed components in this platform that have not been added to the lists of components to be replaced by spares or through cannibalisation still require replacement. When this situation occurs, and the selected platform i cannot be fully restored to the working state, T_{CPSID} and T_{CPCAD} are enabled and fire, removing any tokens that have been added to P_{PCAD} and P_{PSID} placing copies in P_{CAI} and P_{SCJ} respectively, as well as P_{CCRPQ} . This means that the appropriate parts are available in the cannibalisation and spare part inventories once more and that all failed components in this platform are again in the list of components to be replaced.

If tokens representing all failed components can be added to P_{PCAD} and P_{PSID} it means that parts are available to replace all failed components. T_{FCAD} and T_{FSID} fire and pass all tokens in P_{PCAD} and P_{PSID} to P_{CAD} and P_{SID} respectively. These places represent the finalised lists of component replacement decisions. When T_{FCAD} fires, tokens representing the components to be cannibalised are also added to the working component cannibalisation queue, P_{CCAQ} and component replacement queue, P_{CCRPQ} .

When P_{CSPR} is emptied, T_{SPR} is enabled and fires, removing a token from P_{SPR} and placing it in P_{PUM} to indicate that the platform is undergoing maintenance. This enables T_{SRF} as soon as all components are working, as indicated by P_{CW} , and when this transition fires P_{PW} is marked meaning that the platform is returned to the working state.

3.6.3 Component Maintenance Actions

The CMA sub-module, shown in Figure 11, models the replacement of failed components based on component replacement decisions, either by spares or through cannibalisation, and the repair of removed failed components. When a failed platform is added to the platform maintenance queue, its failed components are first removed and then repaired. If a failed component is to be replaced by a spare, maintenance crews install the spare after the failed component is removed. If cannibalisation is to take place, the failed component is replaced by a working component from another failed platform and the maintenance crews remove the failed component and the working component selected for cannibalisation. The meanings of the places and transitions in the CMA sub-module are listed in Table 4.

Transitions T_{RFC} and T_{RWC} represent the removal of failed components from the platform to be repaired and the removal of the required working components from cannibalisation sources respectively. The delay times associated with T_{RFC} and T_{RWC} are obtained by random sampling of the distributions that represent the removal of the appropriate failed and working components respectively. If P_{CRQ} contains a token with colour c , T_{RFC} is enabled and fires after its associated delay has elapsed, removing the token from P_{CRQ} and adding a similar token to both P_{MC} and P_{FCRQ} , which respectively indicate that a particular component is missing from the platform and that the corresponding failed component has been removed and joined a queue to be repaired. The enabling and firing policies of T_{RWC} are similar to those of T_{RFC} except that the input place is P_{CCAQ} , indicating components

that are to be cannibalised, and, in addition to placing a token in P_{MC} to indicate that a component is now missing from the cannibalised platform, a token with colour c is also placed in P_{RWC} , indicating that a working component has been removed from a cannibalisation source and is ready to be placed in another platform.

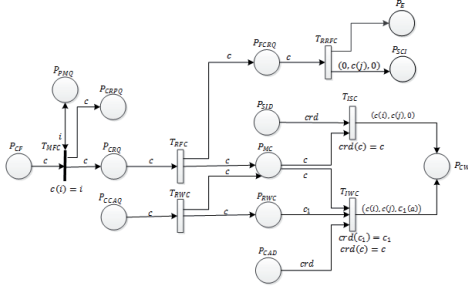


Figure 11. CMA sub-module

Table 4. Places and transitions in the CMA sub-module

Place	Meaning	Transition	Meaning
P_{CRQ}	Component removal queue (components to be removed during maintenance)	T_{MFC}	Maintain failed components
P_{CRPQ}	Component replacement queue (components to be replaced during maintenance)	T_{RFC}	Remove failed components
P_{FCRQ}	Failed components repair queue (failed components that are to be repaired)	T_{RWC}	Remove working components
P_{MC}	Missing components (components that are missing from the system)	T_{ISC}	Install spare components
P_{RWC}	Removed working components (components that are being cannibalised from failed systems)	T_{IWC}	Install working components
P_{CCAQ}	Working component cannibalisation queue (components to be cannibalised from other systems during maintenance)	T_{RRFC}	Repair removed failed components
P_{SID}	Spare installation decision		
P_{CAD}	Cannibalisation decision		
P_{NS}	New spares		
P_E	Event: new spares become available		
P_{SC1}	Spare component inventory		

The transition T_{ISC} represents the installation of spare components based on spare-installation decisions specified by the CRD sub-module. The delay associated with T_{ISC} is randomly sampled from the distribution that represents the installation time of the appropriate spare. T_{ISC} is enabled if places P_{SID} and P_{MC} contain tokens with colour crd and c respectively that satisfy $crd(c) = c$ (meaning that if a component is missing from the platform to be maintained the appropriate spare is selected and installed). T_{ISC} fires after its associated delay has elapsed removing tokens from P_{SID} and P_{MC} and adding a token with the colour $c = (c(i), c(j), 0)$

into place P_{CW} . If there are more than one pair of tokens in the input places of T_{ISC} that satisfy the enabling conditions at the same time, the firing delay of each pair is sampled from a known distribution based on the component type and these firing processes happen concurrently.

The transition T_{IWC} represents installation of working components that have been cannibalised from other platforms, based on cannibalisation decisions specified by the CRD sub-module. The delay associated with T_{IWC} is randomly sampled from the distribution that represents the installation time of the cannibalised component. T_{IWC} is enabled if places P_{CAD} , P_{MC} and P_{RWC} are marked with tokens with colour crd , c and c_1 respectively that satisfy $crd(c) = c$, $crd(c_1) = c_1$ and $c_1(i) \neq 0$, meaning that component c will be replaced by the working component c_1 from the platform $c_1(i)$. The result following the firing of T_{IWC} is similar to that following the firing of T_{ISC} .

T_{RRFC} represents the repair of removed, failed components and has a delay obtained by random sampling of the appropriate component repair time distributions. If P_{FCRQ} contains a token with colour c , T_{RRFC} is enabled and will fire after its associated delay has elapsed, creating a new spare part in P_{SC1} , in the form of a token with colour $c = (0, c(j), 0)$. The availability of new spares may mean that it becomes possible to fully restore failed platforms to the working state, the process of checking if this is possible is conducted in the PRQ sub-module. In order to initiate this process, P_E is also marked after T_{RRFC} fires, which signifies that new spares have become available.

3.6.4 Cannibalisation Inventory Control

The CAIC sub-module, shown in Figure 12, controls the inventory of working components that can be taken from cannibalisation sources to be used as replacements for failed components in other platforms. Since all failed platforms can be used as cannibalisation sources, all working components from failed platforms are added to the component cannibalisation inventory which is a virtual group of the components that are candidates for cannibalisation. To avoid unnecessary cannibalisations, when a platform is selected to be maintained and is able to be fully restored, its working components are removed from the component cannibalisation inventory.

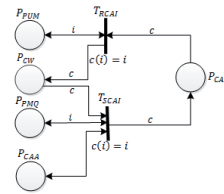


Figure 12. CAIC sub-module

T_{SCAI} ensures that all working components in failed platforms are added to the component cannibalisation inventory and hence made available for cannibalisation. Working components (P_{CW}) from platforms in the maintenance queue (P_{PMQ}) are added to the component cannibalisation inventory (P_{CAI}) if cannibalisation is allowed (P_{CAI} is marked initially).

T_{RCAI} ensures that all working components from platforms that are being restored to the working state are removed from the component cannibalisation inventory. When a platform i is

undergoing maintenance (P_{PUM} contains a token with colour i), any working components from it that are in the cannibalisation inventory (tokens in P_{CAI} with colour c and $c(i) = i$) are moved by the firing of T_{RCM} from P_{CAI} to P_{CWR} .

4. APPLICATION TO A SIMPLE FLEET

4.1 Fleet Description

To demonstrate the application of the CPN model, it is applied to an example fleet studied by Salman et al. [5]. The fleet consists of 10 independent and identical platforms, each of which contains 5 components, which are connected in series and have failure times that follow two-parameter Weibull distributions. The distributions of component failure time, removal/install time and repair time are triangle random variables defined by minimum, maximum and mode values. Maintenance expenditure consists of the costs associated with platform failures, technician wages and spare acquisition and holding costs. The distributions and spare cost parameters are shown in Table 5. Each platform failure is assumed to have a fixed \$150 cost. The manpower hourly wage for removal and installation of components is \$30 while that for repair is \$55.

Table 5. Example fleet parameters [5]

Component	1	2	3	4	5
Life distribution					
Shape parameter	1.0	1.8	2.0	1.0	2.4
Scale parameter	2200	2100	2300	2000	2500
Removal time					
Min(h)	3	5.5	4	2	3
Mode(h)	4	6	5	6	3.5
Max(h)	5	6.5	6	7	4
Repair time					
Min(h)	72	120	60	120	70
Mode(h)	120	124	120	150	100
Max(h)	144	168	140	160	110
Install time					
Min(h)	5	7	6	3	3
Mode(h)	8	10	10	6	4
Max(h)	10	13	12	8	5
Spare acquisition cost (\$)	25000	50000	35000	40000	20000
Annual holding cost (\$)	500	800	700	200	300

The fleet is assumed to perform non-stop (24 hours a day, 7 days a week) missions over a useful life of 5 years. A random selection discipline is applied for platform restoration queuing and the selection of working components that are cannibalised. The CPN model is used to compare the effects of four scenarios that differ according to the use of cannibalisation and the provision of spare components:

- Scenario 1: no cannibalisation, no spares;
- Scenario 2: with cannibalisation, no spares;
- Scenario 3: no cannibalisation, one spare for each component;
- Scenario 4: with cannibalisation, one spare for each component.

4.2 Results

The CPN model was used to carry out 100 simulations of each scenario and the relevant data gathered in each case in order to allow the average fleet performance to be determined. 100 simulations was sufficient to allow the validation of model results

against those presented by Salman et al. [5]. The performance measures computed are the average readiness, which is the average number of functional platforms in the fleet, the average number of failures and the average number of cannibalisations over the fleet's lifetime. The results are summarised in Table 6. The average readiness and average number of failures are in good agreement with those produced by Salman et al. [5]. There are discrepancies in the average number of cannibalisations performed, which are most likely due to differences in how parts are replaced within the models and how spare part availability affects cannibalisation once maintenance is underway.

When no spares are available, performing cannibalisation (scenario 2) brings an increase in the average fleet readiness of approximately 9.97% from 7.590 to 8.347 when compared to the case where no cannibalisation is performed (scenario 1). However, this comes with an increase in costs, with an average 12.58% rise in maintenance costs. As might be expected, since no spares are available cannibalisation is heavily utilised with 48.2% of failures being addressed using cannibalisation. The remaining 51.8% are addressed by component repair.

Table 6. Simulation results of fleet performance

Fleet performance	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Average readiness	7.590	8.347	9.142	9.190
Average number of failures	801.8	875.93	966.39	968.26
Average number of Cannibalisations	0	422.32	0	104.44
Average Cost (\$)	5681717	6396810	7031087	7097011

When there is one spare for each type of component, performing cannibalisation (scenario 4) does not make a large difference to the fleet readiness when compared to the case where no cannibalisation is performed (scenario 3), with a 0.5% increase from 9.142 to 9.190 observed when cannibalisation is introduced. The maintenance cost increases by only 0.94% and 10.79% of component failures are addressed using cannibalisation.

When cannibalisation is not implemented, using spares (scenario 3) provides a substantial increase in average fleet readiness over the case where spares are not used (scenario 1); fleet readiness rises 20.45% from 7.590 to 9.142 and brings an associated 23.74% increase in maintenance expenditure.

When cannibalisation is allowed, using spares (scenario 4) brings an increase in average fleet readiness of 10.1% over the case when there are no spares (scenario 2), from 8.347 to 9.190. There is an associated 10.95% increase in maintenance cost. The availability of spares means that the number of cannibalisations decreases 75.27% from 422.32 per lifetime to 104.44.

The results obtained by applying the CPN model can be used to deduce that, for the modelled fleet:

- Using cannibalisation is beneficial to improve fleet readiness when spare parts are not available;
- Using spare parts can bring a substantial benefit to fleet readiness.
- It is better not to perform cannibalisation when spares are available.

5. SUMMARY AND CONCLUSIONS

A CPN model of fleet cannibalisation has been developed. A number of operational and maintenance-related factors are considered including fleet operation, platform failure logic, platform restoration queuing, cannibalisation, spares and repair. The CPN model is modular, consisting of fleet, platform, component and maintenance modules. In order to demonstrate the application of the model, an example fleet with fixed missions was simulated. The results were used to analyse the effect of cannibalisation on the fleet's performance, using a number of performance measures. For the studied fleet, the results obtained demonstrated that cannibalisation could improve fleet readiness and also that both using cannibalisation and keeping a spare part inventory gave the best possible fleet readiness.

The CPN model presented in this paper brings with it a number of benefits for fleet maintenance and cannibalisation modelling:

- The model can be easily adapted to investigate other platforms and fleets;
- The size of the model does not increase when larger fleets are considered since the number of platforms in a fleet and the number of components in a platform are changed simply by changing the initial marking of the relevant places;
- Since the CPN model is modular it can be easily changed to model different scenarios or extended in order to consider other factors related to maintenance.
- The model can be easily used to gather other information about the effects of maintenance and the spare part inventory on fleet operation and performance.

6. ACKNOWLEDGEMENTS

Darren Prescott is the Lloyd's Register Foundation (LRF) Lecturer in Risk and Reliability, based in The LRF Centre for Risk and Reliability Engineering at the University of Nottingham and would like to gratefully acknowledge the support of the LRF.

Lloyd's Register Foundation (LRF), a UK registered charity and sole shareholder of Lloyd's Register Group Ltd, invests in science, engineering and technology for public benefit, worldwide.

7. REFERENCES

- [1]. Fisher, W.W. (1990). "Issues and Models in Maintenance Systems in Incorporating Cannibalisation: a Review". *INFOR: Information Systems and Operational Research*. 28(2), pp.67-88.
- [2]. Fisher, W.W. (1990). "Markov Process Modeling of a Maintenance System with Spares, Repair, Cannibalization and Manpower Constraints". *Mathematical and Computer Modelling*. 13(7), pp. 119-125.
- [3]. Cassady, C. R., Mason, S.J. and Yeung, T. G. (2003). "Quantifying the Impact of System Cannibalization". Research Report. U.S. Air Force Research Laboratory.
- [4]. Ormon, S.W. and Cassady, C. R. (2004). "Cannibalization Policies for a Set of Parallel Machines". In *Reliability and Maintainability, 2004 Annual Symposium*.
- [5]. Salman, S., Cassady, C. R., Pohl, E.A., and Ormon, S.W. (2007). "Evaluating the impact of cannibalization on fleet performance". *Quality and Reliability Engineering International*. 23(4), pp. 445-457.
- [6]. Jensen, K. and Kristensen, L.M. (2009), *Coloured Petri Nets: Modelling and Validation of Concurrent Systems*, Springer, Heidelberg.
- [7]. Petri, C.A. (1962), *Kommunikation mit automaten*, *PhD Dissertation*, Darmstadt University of Technology

Attribute Control Chart Development by Evidential Reasoning

Fattaneh Javadi
MSc Industrial Engineering
Islamic Azad University
Karaj Branch, Iran
+(98)9125637090
fattaneh.javaadi@gmail.com

Farzaneh Ahmadzadeh
Post Doc Research assistant
Division of Product realization
Malardalen university, Sweden
+(46)702460950
farzaneh.Ahmadzadeh@mdh.se

Abolfazl Mirzazadeh
Associate Professor
Dept of Industrial Engineering
University of Kharazmi, Tehran, Iran
+(98)2144451990
a.mirzazadeh@aut.ac.ir

ABSTRACT

Attributes charts are commonly used in monitoring quality characteristics of the proportion type and these charts assume that the monitored characteristics are binomially distributed. Classical control charts need to certain and precise data. However, in practice, quality experts express their opinion in imprecise form, which in turn, add more uncertainty and ambiguity. It is essential to properly represent and interpret uncertain information to evaluate product items. In this paper, the evidential reasoning (ER) based approach has been developed for supporting this uncertainty. So the belief multinomial p-control chart is introduced for monitoring the production process in the uncertainty condition, using evidence theory and the belief structure. A numerical example showing production process evaluation is examined by using the ER approach. The results show the proposed approach, is effective not only for reducing production defective but also for increasing the certainty in interpreting of quality variables (data).

Keywords

Belief multinomial control chart, P-control chart, Evidence theory, Uncertainty condition, Linguistic variable, Belief Structure.

1. INTRODUCTION

Control charts are important tools of statistical quality control, which aims to achieve a desired level of confidence, ease of judgment and development of qualitative studies. The power of control charts lies in their ability to detect process shifts and to identify abnormal conditions in the process.

Walter Shewhart (1924) designed the first control chart. He proposed a general model for control charts as follows; Let w be a sample statistic that measures some quality characteristic of interest. Moreover, suppose that the mean of w is μ_w and the standard deviation is σ_w . Then the center line (CL), the upper control limit (UCL), and the lower control limit (LCL) are defined as follows:

$$\begin{aligned} UCL &= \mu_w + 3\sigma_w \\ CL &= \mu_w \\ LCL &= \mu_w - 3\sigma_w \end{aligned} \quad (1)$$

A single measurable quality characteristic such as dimension, weight or volume is called a variable. In such cases, control charts for variables such as \bar{x} -chart and R-chart (or S-chart) are used. If the quality-related characteristics such as appearance, softness,

colour, taste, etc., attributes control charts such as p-chart, c-chart are used to monitor the production process. Sometimes the product units are classified as either "conforming" or "nonconforming", depending upon whether or not they meet some specifications. The p-chart is used to monitor the process based upon the fraction of nonconforming units.

Linguistic scales are commonly used in industry to express characteristic of products. Quite often, expert team express their opinion estimated and imprecise. In such human judgments, linguistic variables are used to express the assessments, and a linguistic variable differs from a numerical variable in that its values are not numbers but words or phrases in some language. In monitoring vague production processes, (Wang et al., 1990; Raz et al., 1990; Kanagawa et al., 1993) developed control charts for linguistic variables where the vague observations can be recorded to exact numbers, which can be used to establish classical control charts. When the products are classified into mutually exclusive linguistic categories, fuzzy control charts are used frequently and different procedures are proposed to construct these charts. Fuzzy Multinomial chart with the fixed sample size (Amirzadeh et al., 2008) and Fuzzy Multinomial process with Variable Sample Size (Pandurangan et al., 2011) have developed.

Dempster-Shafer (D-S) theory is based on the work of Dempster during the 1960's and of Shafer during 1970's (Shafer 1976; Dempster et al., 2008). The Evidential Reasoning (ER) approach is developed on the basis of D-S theory (1976) and decision theory. By introducing the concepts of belief structure (Yang et al., 2002), for the first time it becomes possible to model uncertainties of various types of nature in a unified format for further analysis without resorting to sensitivity analysis. Since the introduction of the modelling technique using belief structure (Zhang et al. 1990) and the development of the ER approach (Yang et al. 1994), significant amounts of work using this approach have emerged in literature, including; Location Problem (Rahgan et al. 2012), Risk Evaluation (Gao et al. 2010), Expert Systems (Neumann et al., 2012), Multi criteria Decision analysis (Xu, 2012), Suppliers Prioritization (Massahi et al. 2012).

Moreover, often by using non-fuzzy operators, too much information is lost. Moreover, different operators lead to different decisions about which processes, that these are the reason of lack of efficiency for these methods. Also, quality experts express their opinion in imprecise form. In this paper, a multinomial belief P-control chart (BP chart) for linguistic variable is proposed, using evidential reasoning and belief structure. The BP-chart deals with a linguistic variable, which is classified into more than two categories.

2. Belief P-control chart

Based upon evidence theory, an evaluation grades is defined by a set on all of the quality characteristics for evaluating production items, as:

$$H = \{H_1, H_2, \dots, H_N\} \quad (2)$$

For example, an evaluation and visual control of the production items might have the following assessment possibilities:

- "Excellent conforming", if the product works and has neither defects nor aesthetic flaws of any kind.
- "Good conforming", if the product works and has no defects, but it has some aesthetic flaws.
- "Medium conforming", if the product works but has some defects.
- "Poor conforming", if the product works and has no defects, but has only a few aesthetic flaws.
- "Nonconforming", if the product does not work correctly.

So, the set of evaluation grades will be: $H = \{\text{Excellent conforming } (H_5), \text{ Good conforming } (H_4), \text{ Medium conforming } (H_3), \text{ Poor conforming } (H_2), \text{ Nonconforming } (H_1)\}$. In evaluation of qualitative attributes, uncertain judgment can used.

In this section, a new approach for construction of BP-chart is proposed. The statistical principles underlying the multinomial Belief P-control chart (BP-chart) are based on the multinomial distribution. As defined in (2), H is a set of evaluation grades, which can take N mutually exclusive members. Suppose that the production process is operating in a stable manner, and P_i is the probability that an item is evaluated in H_i , $i = 1, 2, \dots, N$. Moreover, successive items produced are independent. Assume that a random sample of n products is selected and let X_i , $i = 1, 2, \dots, N$, be the number of items of the product that are H_i , $i = 1, 2, \dots, N$. Then (X_1, X_2, \dots, X_N) has a multinomial distribution with parameters n and (p_1, p_2, \dots, p_N) . It is known that each X_i , $i = 1, 2, \dots, N$, marginally has a binomial distribution with the mean np_i and variance $np_i(1-p_i)$, $i = 1, \dots, N$, respectively. P_i is calculated by the following:

$$P_i = \frac{D_i}{n} \quad (3)$$

Assume that, there is k quality experts with relative importance $\lambda_k > 0$, $k=1, \dots, K$, which $\sum_{k=1}^K \lambda_k = 1$, for evaluating the product items. A given assessment for an item may mathematically defined by the following distribution:

$$\{(H_{ij}, \beta_{ij}^{(k)}), i = 1, \dots, N; j = 1, \dots, N\} \quad (4)$$

Where $\beta_{ij}^{(k)} \geq 0$, $\sum_{n=1}^N \beta_{ij}^{(k)} \leq 1$, and $\beta_{ij}^{(k)}$ denotes a degree of belief that is the assessment of k^{th} expert ($k=1, \dots, K$). The above distributed assessment reads that the item is assessed to the grade H_{ij} with the degree of belief $\beta_{ij}^{(k)}$, $i=1, \dots, N$; $j=1, \dots, N$. Quality experts assign each item to evaluation grades and also assign the

degree of belief β_{ij} to each item. Then, each quality expert expected score is calculated by the following:

$$E(S_i^{(k)}) = \sum_{n=1}^N H_{ij} \beta_{ij}^{(k)} \quad (5)$$

And then, group belief degree of team members would have the following belief structure:

$$\{(H_{ij}, \beta_{ij}), i = 0, \dots, N; j = 1, \dots, N\} \quad (6)$$

Which

$$\beta_{ij} = \sum_{k=1}^K \lambda_k \beta_{ij}^{(k)}, i=1, \dots, N; j=1, \dots, N \quad (7)$$

And β_{ij} is the belief degree of H_{ij} interval, with $\sum_{i=1}^N \sum_{j=1}^N \beta_{ij} \leq 1$. Aggregated expected score of K experts for each item is calculated by

$$E(S_i) = \sum_{k=1}^K \lambda_k E(S_i^{(k)}) \quad (8)$$

(Chin et al., 2009). Now the number of items belong to each evaluation grade and also the proportion of these items is determined by the results of equation (8). Note that, for determining the evaluation grade of itmes that have point assessment but non-integer results, following rules are defined:

1. For measures between 1 and 2; if the result is less than 1.5 the item would belong to H_1 evaluation grade, and if it is equal or greater than 1,5 the item would belong to H_2 .
2. For measures between 2 and 3; if the result is less than 2.5 the item would belong to H_2 evaluation grade, and if it is equal or greater than 2,5 the item would belong to H_3 .

And so on. Then, a total belief degree is assigned to each item of i^{th} category in each sample, denoted by \bar{S}_i using:

$$\left\{ \left(H_{ij}, \frac{1}{n} \sum_{l=1}^L \beta_{ijl} \right), i = 1, \dots, N; j = 1, \dots, N \right\} \quad (9)$$

The weighted average of linguistic variable, denoted by \bar{S} is defined by:

$$\bar{S} = \frac{\sum_{i=1}^N x_i \bar{S}_i}{\sum_{i=1}^N X_i} = \sum_{i=1}^N p_i \bar{S}_i \quad (10)$$

We introduce the control limits for BP-chart as:

$$UCL = E(\bar{S}_i) + 3\sqrt{Var(\bar{S}_i)} \quad (11)$$

$$CL = E(\bar{S}_i)$$

$$LCL = E(\bar{S}_i) - 3\sqrt{Var(\bar{S}_i)}$$

that $E(\bar{S}_i)$ and $Var(\bar{S}_i)$ for each sample is computed by the following equations:

$$E(\bar{S}_i) = \sum_{i=1}^N p_i \bar{S}_i \quad (12)$$

$$Var(\bar{S}_i) = \frac{1}{n} \left[\sum_{i=1}^N \bar{S}_i^2 p_i (1-p_i) - 2 \sum_{i=1}^N \sum_{j=1}^N \bar{S}_i \bar{S}_j p_i p_j \right] \quad (13)$$

3. Numerical Study

In a gas water heater manufacturing process, the P-chart is used for assess the production process. Expert team express their opinion estimated and imprecise, and linguistic variables are used to express the assessments. Production items may be classified by an expert team as either "Excellent conforming", "Good conforming", "Medium conforming", "Poor conforming" or "Nonconforming". For this study, 2 samples of size 10 are selected in a week during 10 weeks. Also the expert team includes 3 quality experts with $\lambda_1=\lambda_2=\lambda_3=1/3$. The data of the assessments are given in Table-1. A multinomial belief P-control chart (BP chart) for linguistic variable is proposed, using evidential reasoning and belief structure. As we said, the BP-chart deals with a linguistic variable which is classified into more than two categories. Of course, only data and results of calculation of 10 items of sample 1 are shown in the related tables.

Table 1. Data of the sample products assessment

Sample No	Item No	Q1	Q2	Q3
1	1	5	4-5:1	(5,0.5)(4,0.5)
	2	(4,0.9)(5,0.1)	3-4:0.9	3-4:1
	3	4	5	4-5:1
	4	5	(5,0.95)(4,0.05)	4-5:0.9, 3,0.1
	5	4	(4,0.7)(5,0.2)	4-5:1
	6	3	(3,0.6)(4,0.4)	(3,0.5)(4,0.5)
	7	5	(5,0.8)(4,0.2)	4-5:1
	8	2	(2,0.4)(3,0.6)	3
	9	4	unknown	5
	10	1	1	1

Both Point and interval assessment are (Wang, 2013) given in the above table. Then, group belief degrees and expected scores of each of expert are calculated by (5) to (7), and given in Table-2.

Table 2. Group belief product items and expert's Expected Scores

Sample No	Item No	Group Belief	Q1	Q2	Q3
1	1	(5, 0.5)(4,0.17)(4-5,0.33)	5	4-5	4.5
	2	(5,0.033)(4,0.3)(3-4,0.63)	4.1	2.8-4.1	3-4
	3	(5,0.33)(4,0.33)(4-5,0.33)	4	5	4-5
	4	(5,0.65)(4,0.017)(3,0.033)(4-5,0.3)	5	4.95	3.9-4.8
	5	(5,0.067)(4,0.57)(1-5,0.033)(4-5,0.33)	4	3.8	4-5
	6	(4,0.3)(3,0.7)	3	3.4	3.5
	7	(5,0.6)(4,0.067)(4-5,0.33)	5	4.8	4-5
	8	(3,0.53)(2,0.47)	2	2.6	3
	9	(5,0.33)(4,0.33)(1-5,0.33)	4	1-5	5
	10	(1,1)	1	1	1

And the expected score of all quality experts of each item are calculated by (8). These amounts are given in Table-3. And evaluation grades related to each item are shown in column 4 of Table-3. So the number of items that are H_i are obtained is determined. The value of P_i are given in Table-4. Then, the total belief degree of each sample is calculated by (9) and are given in Table-5.

Table 3. Total expected scores of 3 experts for each product item

Sample No	Item No	$E(S_i)$	Evaluation Grades related
1	1	4.5-4.83	4--5
	2	3.3-4.07	3--4
	3	4.33-4.67	4--5
	4	4.62-4.92	4--5
	5	3.93-4.27	3--5
	6	3.3	3
	7	4.6-4.93	4--5
	8	2.53	3
	9	3.33-4.67	3--5
	10	1	1

Table 4. Probability of each category in each sample

Evaluation Grades Item No	1	1-2	1-3	1-4	1-5	2	2-3	2-4	2-5	3	3-4	3-5	4	4-5	5
1	0.1	0	0	0	0	0	0	0	0	0.2	0.1	0.2	0	0.4	0

Table 5. Total group belief

Evaluation Grades Item No	1	1-2	1-3	1-4	1-5	2	2-3	2-4	2-5	3	3-4	3-5	4	4-5	5
1	0.1	0	0	0	0.04	0.05	0	0	0	0.13	0.06	0	0.21	0.163	0.252

Now, the weighted average of linguistic variable, \bar{S}_i , illustrated in table-6 are calculated by (9).

Table 6. Weighted average of linguistic variables

Sample NO	1	2	3	4	5	6	7	8	9	10
\bar{S}_i	0.107	0.08767	0.05333	0.5167	0.1227	0.0898	0.08833	0.06733	0.04567	0.099
Sample No	11	12	13	14	15	16	17	18	19	20
\bar{S}_i	0.09933	0.14	0.056	0.06	0.7967	0.11283	0.11567	0.08233	0.065	0.10667

$P_{Non-Poor}=0.08$, $P_{Poor}=0.03$, $P_{Medium}=0.075$, $P_{Good}=0.135$, $P_{Excellent}=0.08$, $P_{Non-Poor}=0.1$, $P_{Non-Medium}=P_{Non-Good}=0$, $P_{Non-Excellent}=0.005$, $P_{Poor-Medium}=0.02$, $P_{Poor-Good}=0.015$, $P_{Poor-Excellent}=0.005$, $P_{Medium-Good}=0.11$, $P_{Medium-Excellent}=0.1$, $P_{Good-Excellent}=0.245$. Hence, the central line and control limits are: CL = 0.10883, UCL = 0.19442, LCL = 0.02324.

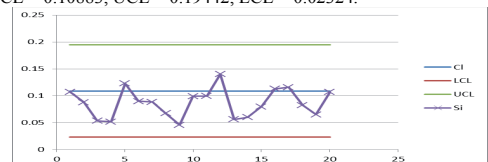


Figure 1. BP-chart of the \bar{S}_i for 20 samples

All of the samples are in the control limits. So the process is in-control. And with acceptance of control limit, step 1 for estimating control limits is finished. Now, the \bar{S}_i for the 20 samples are plotted in the BP-chart (Figure 1). In the conventional p-charts, the center line represents the fraction of nonconforming items. Therefore the central line and control limits for the p-chart are: CL = 0.13, UCL = 0.2013, LCL = 0.0586. The P-chart for the 20 samples is shown in Figure 2.

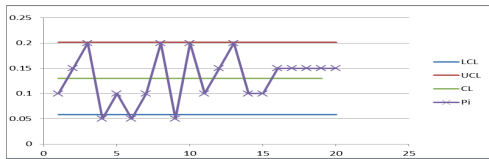


Figure 2. P- chart of the 20 samples

Some of the samples like 4, 6, 9 are outside of the control limits that illustrates the process is out-of-control. But considering this samples and finding the cause deviations and calculating the new control limits are not in the scope of this paper.

4. Conclusion

The theory of classical control charts requires all the data to be exactly known. But the data to construct attributes control charts includes human subjectivity and vagueness. So data are ambiguous or not well defined. In this paper, a platform to develop quality control charts literature is established with aiming to extend the statistical quality control methodology which deals with uncertainty underlying the manufacturing processes. So evidential reasoning is used. BP-chart has been proposed for linguistic data set, aim to reduction of uncertainty. To draw the chart, samples of fixed sample size are chosen randomly from a production line. For illustrating the efficiency of proposed approach, and accuracy in this approach, a numerical example is used. Also in the example more than two categories are used for evaluating items and the number of evaluation grades is 5 categories and 10 interval evaluation grades between them. The new chart is more efficient than the classical p-chart. Interested people can develop the proposed approach for monitoring multivariate control chart, or extend and develop proposed approach for other types of attribute control charts such NP- chart and C-chart.

5. REFERENCES

- [1] Amirzadeh, V. Mashinchi, M. and Yaghoobi, M.A. (2008). "Construction of Control Charts Using Fuzzy Multinomial Quality". *Journal of Mathematics and Statistics*. Vol. 4, No. 1. pp. 26-31.
- [2] Chin, K. S. Wang, Y. M. Yang, J. B. & Poon, K. K. G. (2009). "An evidential reasoning based approach for Quality function Deployment under uncertainty". *Expert Systems with Applications*. Vol. 36. 5684-5694.
- [3] Dempster, A. P. Yager, R. R. & Liu, L. (2008). *The classic works on the Dempster-Shafer Theory of Belief Functions*. Springer-Verlag.
- [4] Dezert, J. Wang, P. & Tchamova, A. (2012). "On the validity of Dempster-Shafer Theory". 15th International Conference on Information Fusion, Singapore.
- [5] Gao, L. Mock, T. & Srivastava, R. P. (2010). "An Evidential Reasoning Approach to Fraud Risk Assessment under Dempster-Shafer Theory: A General Framework". Working Paper, School of Business, The University of Kansas.
- [6] Kanagawa, A. Tamaki, F. & Ohta, H. (1993). "Control charts for process average and variability based on linguistic data". *International Journal of Production Research*. Vol. 31, No. 4, pp. 913-922.
- [7] Massahi, M. Mirzazadeh, A. & Mirzadeh, P. (2012). "The Evidential Reasoning Approach for Suppliers Prioritization with Group-AHP Weights". Proceedings of the 2012 International Conference on Industrial Engineering and Operations Management Istanbul, Turkey, July 3 – 6, pp. 1945-1953.
- [8] Neumann, T. & Weinrit, A. (2012). "Expert systems based on Dempster-shafer theory of evidence in maritime-characteristic and applications". *Scientific Journals*. Vol. 32, No. 104. pp. 141-147.
- [9] Pandurangan, A. & Varadharajan, R. (2011). "Fuzzy multinomial Control chart with variable sample size". *International Journal of Engineering Science and Technology*. Vol. 3, No. 9, pp. 6984-6991.
- [10] Rahgan, Sh. & Mirzazadeh, A. (2012). "A new method in the location Problem using Fuzzy Evidential Reasoning". *Research Journal of applied Sciences, Engineering and Technology*. Vol. 4, No. 22, pp. 4636-4645.
- [11] Raz, T. & Wang, J. H. (1990). Probabilistic and membership approaches in the construction of control charts for linguistic data, *Production Planning & Control*. Vol. 1, No. 3, pp. 147-157.
- [12] Shafer, G. A. (1976). *Mathematical theory of evidence*. Princeton: Princeton University Press.
- [13] Wang, J. H. & Raz, T. (1990). "On the construction of control charts using linguistic variables". *International Journal of Production Research*. Vol. 28, No. 3, pp. 477-487.
- [14] Wang, J. Q. (2013). "Intuitionistic fuzzy multi-criteria decision-making method based on evidential reasoning". *Applied Soft Computing*. Vol. 13, No. 4, pp. 1823-1831.
- [15] Xu, D. L. (2012). "An introduction and survey of the evidential reasoning approach for multiple criteria decision analysis". *Annals of Operations Research*. Vol. 195, No. 1. pp. 163-187.
- [16] Yang, J. B. & Singh, M. G. (1994). "An evidential reasoning approach for multiple attribute decision making with uncertainty". *IEEE Transactions on Systems, Man, and Cybernetics*. Vol. 24, No. 1. pp. 1-18.
- [17] Yang, J. B. & Xu, D. L. (2002). "On the evidential reasoning algorithm for multi attribute decision analysis under uncertainty". *IEEE Transactions on Systems, Man and Cybernetics*. Part A. Systems and Humans. Vol. 32, No. 3. pp. 289-304.
- [18] Zhang, Z. J. Yang, J. B. & Xu, D. L. (1990). "A hierarchical analysis model for multi objective decision making, in analysis, Design and Evaluation of Man-Machine Systems". Selected papers from the 4th IFAC/IFORS/IEA, Xian, P. R. china, September 1989. Pergamon, Oxford, UK, 13-18.

Detection and Diagnosis of Broken Rotor Bar based on the Analysis of Signals from a Variable Speed Drive

<p>D. Ashari Centre of Efficiency and Performance Eng. University of Huddersfield Queensgate Huddersfield HD1 3DH, UK +44 01484 471193 djoni.ashari@hud.ac.uk</p>	<p>M. Lane Centre of Efficiency and Performance Eng. University of Huddersfield Queensgate Huddersfield HD1 3DH, UK mark.lane@hud.ac.uk</p>	<p>F. Gu Centre of Efficiency and Performance Eng. University of Huddersfield Queensgate Huddersfield HD1 3DH f.gu@hud.ac.uk</p>	<p>A.D. Ball Centre of Efficiency and Performance Eng. University of Huddersfield Queensgate Huddersfield HD1 3DH, UK a.d.ball@hud.ac.uk</p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------

ABSTRACT

In this paper, the diagnosis of broken rotor bar (BRB) is investigated analytically and experimentally based on induction motors (IM) with variable speed drives (VSD). The analysis of the VSD process has understood that the sensorless control mode adjusts both supply voltage and current in order to maintain the operating speed within high accuracy. It means that any asymmetric problems with the rotor such as BRB will cause changes to current, voltage and power signals. Therefore, all of these signals can be used to diagnose BRB. However, experimental studies show that the components at twice slip frequency can be extracted more reliably in the spectrum of power signal, which contains diagnostic information of both current and voltage signals. Consequently, it produces better results in separating different BRB cases under sensorless control mode, compared with that of using either current or voltage signal.

Keywords

Broken rotor bar, Variable speed drive, Motor current signature analysis, Instantaneous power,

1. INTRODUCTION

Broken rotor bars (BRB) in induction motors are a common fault which often brings low efficiency operation and even unexpected breakdowns, leading to loss of productivity. To provide timely warning of the fault, many studies have been carried out to develop more accurate diagnostic methods. As shown in the review papers [1, 2], most previous works utilises the spectrum analysis of phase current signals from which the sideband components at frequency $(1 \pm 2ks)f_s$ (s is the rotor slip and f_s is the supply frequency, and harmonic integer ($k = 1, 2, 3, \dots, n$) are extracted to diagnose such faults.

However, these studies have been carried out based on motors without modern variable speed drives (VSD), which have been widely used in many plants to achieve high efficient production. In general, a VSD can be configured to operate under with a sensorless control mode or a volt/Hz mode. Both of these modes adjust both supply voltage and current in order to maintain the operating speed within high accuracy. It means that the effect of BRB will reflect on both of these signals, which is significantly different from motors without VSD where only current signal reflects BRB features.

Therefore, this studies focus on evaluating the diagnostic features of BRB cases under sensorless control mode using current, voltage and power signals. To understand of what the results of this case, the volt/Hz mode is analysed and presented as a comparison purposes.

Different analysis methods such as spectrum have been applied to stator current signals, which is known as motor current signature analysis (MCSA) and employed as a popular technique for BRB diagnosis. In the meantime, instantaneous power (IP) spectrum also becomes an effective monitor of machine health in order to remove the effect of the high amplitude at supply frequency and simplify the complexities of the stator current and axial flux spectra. The work was started [3] by showing that in the current spectrum shows two sideband components whereas in IP spectrum just show the low frequency component due to the fault on rotor. This shows the superiority of IP to the current stator spectra shown on (1)[3].

$$\begin{aligned}
 P_{AB}(t) &= P_{AB,0}(t) + \frac{1}{2} MV_{LL} I_L \left\{ \cos\left[(2\omega + \omega_{OSC})t - \varphi - \frac{\pi}{6}\right] \right. \\
 &\quad \left. + \cos\left[(2\omega - \omega_{OSC})t - \varphi - \frac{\pi}{6}\right] \right. \\
 &\quad \left. + 2 \cos\left(\varphi + \frac{\pi}{6}\right) \cos(\omega_{OSC}t) \right\} \\
 P_{CB}(t) &= P_{CB,0}(t) + \frac{1}{2} MV_{LL} I_L \left\{ \cos\left[(2\omega + \omega_{OSC})t - \varphi + \frac{5\pi}{6}\right] \right. \\
 &\quad \left. + \cos\left[(2\omega - \omega_{OSC})t - \varphi + \frac{5\pi}{6}\right] \right. \\
 &\quad \left. + 2 \cos\left(\varphi - \frac{\pi}{6}\right) \cos(\omega_{OSC}t) \right\} \\
 P_{ABC}(t) &= P_{ABC,0} [1 + M \cos(\omega_{OSC}t)]
 \end{aligned} \tag{1}$$

Where V_{LL} , I_L are rms values of line-to-line voltage and line currents, respectively; ω is supply frequency in radians per second, φ is power factor angle and M is modulation depth.

The last technique is shaft voltage or current monitoring. Many electrical utilities have tried to monitor voltage in the hope that they may be an indicator of core or winding degradation, because they give rise to large shaft currents. However, the shaft voltage on its own gives no useful parameter for monitoring as the voltage is difficult to measure continuously and the damage to the machine needs to be substantial before a significant variation in shaft voltage occurs [4]. However, these voltage parameters become useful when these are used with the presence of the current parameters, as IP.

It has found that there are limited studies on VSD based BRB detection and diagnosis. Especially, at comparison with different

signals, operation modes: volt/Hz and sensorless mode. The paper will examine the detection and diagnosis performance of current, voltage and instantaneous power by using spectrum analysis. A 4-kW three-phase squirrel-cage IM and voltage source inverter with vector-control technique for open loop and sensorless controlled were used for the experiment. Detailed of the IM specifications can be seen on the Appendix and operating principle of the voltage source inverter is discussed later on the next section.

The paper has further five more sections. Section 2 explains the principle of VFD and how BRB influences current and voltage in a VSD system. Section 3 describes the test system and test procedure. Section 4 discusses the results and Section 5 is the conclusion.

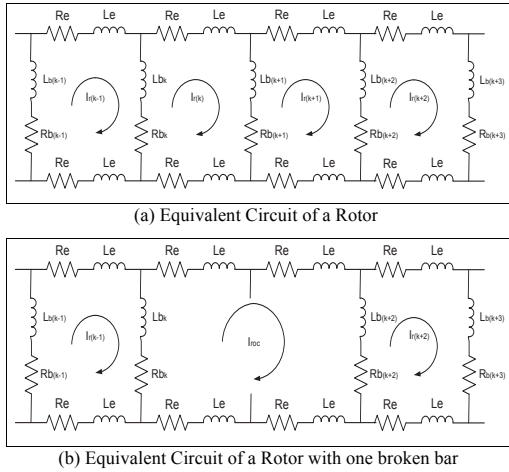


Figure 1. Equivalent Circuit of a Rotor with one broken bar

2. PRINCIPLES OF VSD AND HOW BRB INFLUENCES CURRENT AND VOLTAGE IN A VSD SYSTEM

2.1 Control Process

Extensive VSD control has been reported [5], basically developed for the understanding of ac motor equivalent circuit, then Clarke and Park Transforms and concept on direct-quadrature (d-q) control. The d-axis and q-axis are perpendicular and d-axis represents the speed and q-axis represents torque of the motor to be controlled. The so called FOC consists of direct and indirect/sensorless controller. Direct Field Orientation Control (DFOC) which estimates the angle of machine by making use of the sensor of machine or an observer which makes use of the measured electrical quantities at terminals of the machines. This term called sensorless. Indirect Field Oriented Control (IFO) uses mechanical sensor/encoder to measure the shaft speed.

Two control modes for Field Oriented Control (FOC) are torque and flux linkage control. FOC allows independent control of torque and flux linkage under transient condition. For the speed

control mode, the torque feed forward is embedded for calculation of the motor electromagnetic torque.

The speed observer structure for calculating the rotor speed in the absent of an encoder is included in the controller. The Model-reference Adaptive System (MARS) is used for the drive controller. MARS estimates the speed to improve the accuracy of rotor flux estimation.

The speed observer structure and its frequency characteristics should be aware of:

- controllers parameters tuning;
- state variables estimators parameters tuning; and
- machine equivalent circuit parameters estimation accuracy

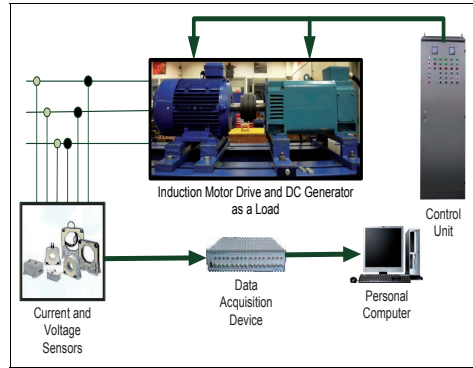


Figure 2. Schematic diagram of the induction motor test facility

For a high-performance drive non-linear system identification of a motor drive system is required and the results of system identification could provide the controller with information regarding the load variation, system noise, and parameter variation of the induction [6]. This is needed for the controller parameter tuning.

2.2 BRB induced changes in current and voltage

Broken rotor bar and end-ring condition is one of induction machine asymmetrical stator and/or rotor winding connection. This asymmetrical operation results in unbalanced air gap voltage, consequently unbalanced line currents, increased losses, increased torque pulsations and decrease average torque. At the most severe faults, it will result in poor efficiency and excessive heating which eventually leads to the failure of the machine.

Figure 1(a)-(b) show the equivalent circuits of the BRB showing the rotor loop current and end ring current. Detailed calculations are proposed in [7]. The results are that marked increases in the lower sideband of the first harmonic (LSB1), which is $(1 - 2s)$ times the synchronous frequency, corresponding to $(1 \pm 2ks)f_s$ in current spectrum.

BRB can be explained further [8] by suggesting that sensors should be non-invasive, reliable, accurate diagnosis, severity of

the problem should be quantified, ideally an estimation of remaining life time and on-line. Quantification of the severity of faults and remaining lifetime are topics to be explored into details on the future works.



Figure 3a. Half Broken Rotor Bar configuration



Figure 3b. One Broken Rotor Bar configuration

More details on current signature analysis have been presented in [9] on the new improved BRB diagnosis based on modulation signal bispectrum analysis. The current signals on the healthy machine and under BRB are explained. The ideal electromagnetic relationship of the driving motor on both healthy and under BRB can be examined. The results are that the additional current with sinusoidal wave at frequency $2sf_s$ for the case of BRB can be found. This so called fault frequency for the case of BRB has components of amplitude and phase angle. It is shown that current signal of the faulty case exhibits three new additional components compared to the healthy one. Two of these are the lower sideband components at the same frequency, but with different phases, and the other is the upper sideband component with a phase different from the previous two.

3. TEST SYSTEM AND TEST PROCEDURE DESCRIPTION

The test motor is a three-phase IM with rated output power of 4 kW at speed 1420 rpm (two-pole pairs). To change the speed of the motor, a digital variable speed controller is attached to the test rig between the power line source and the motor. The controller

can be programmed to any specific shaft rotation speed. The IM is directly coupled with a loading DC generator. The field of the generator is connected to DC source through controller while the generated power was fed back to the mains electrical grid and the load in the IM can be adjusted by changing the field resistance of the DC generator. The operating load can be varied from no load to full load via the control panel.

Figure 2 shows schematic diagram of the IM test facility. The system consists of an IM, variable speed controller, coupling and DC generator as a load.

A power supply measurement box was designed to measure AC voltages and currents, using Hall Effect voltage and current transducers.

During the experimental work all the data was acquired using a YE6232B data acquisition system. This system has 16 channels, each channel with a 24 bit analogue-digital converter with a maximum sampling frequency of 96 kHz.

The voltage signals produced by inverter in the form of pulse width modulation (PWM) were filtered and put on Hilbert modulation converter.

One healthy and two predefined faults of IMs are tested. To represent faults between the case of short and open circuits, two faults of the motors are half broken on (Fig. 2) and one broken rotor bar on (Fig. 3) faults. IMs are identical with 4kW, Δ -connection and called healthy, half and one BRB. The data sets taken out are to be rotor speed, rms calculation from 3-phase currents and voltages, and spectrum calculation. Faults were induced by drilling carefully into the bars along their height in such a way that the hole cut the bar completely to simulate the broken rotor bar fault and then compared with a healthy motor under full constant speed and five different torque loads (0%, 25%, 50%, 75% and 100%). Each test acquired data at thirty seconds record and sampling rate of 96 kHz.

Once the data have been analysed, basic operating parameters were displayed. The operating parameters are speed, current and voltage. Some basic parameters for data acquisition were set, such as sensitivity of the current and voltage. Current spectra were analysed to see if any changes due to supply unbalance. The temperature sensor was used to monitor the temperature in specific normal operating condition.

When checking of the basic operating parameters is done, the next stage is calculating efficiency estimation, which is the electrical power as input and the shaft power as output. The last stage is to analyse the spectrum characteristics. The comparison between the current, voltage and instantaneous spectra is carried out for both open loop and sensorless control methods on the amplitudes of the slip and twice-slip frequencies.

4. RESULTS AND DISCUSSION

To obtain the BRB component, a conventional spectrum is applied to the current, voltage and IP signals in order to evaluate their performance respectively. The spectrum is calculated using FFT from the time domain data of six repeating tests. The size of FFT is 524288, leading to a frequency resolution of 0.1831Hz because of sampling rate is 96 kHz; averaged up to 78 times. To reduce spectral leakage, a Hanning data window is applied to the data frame.

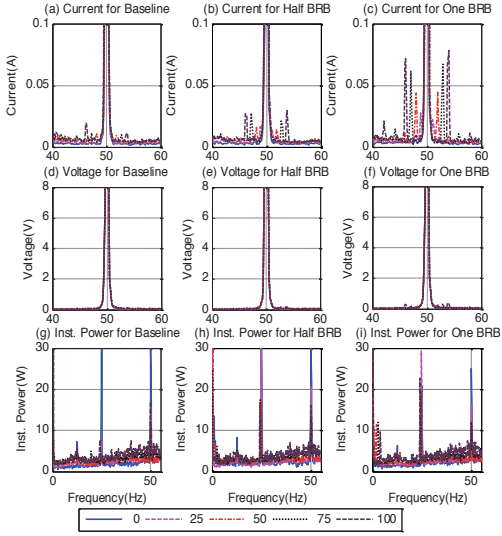


Figure 4. Spectra of current, voltage and IP for different loads under volt/Hz operation

The high frequency resolution of 0.1831 Hz is needed to make a reliable identification of the sideband amplitudes. For low load operation, changes of the slip are considerably small (less than 0.02) which result in a sidebands located about less than 1Hz around the supply frequency..

4.1 Spectra of Current, Voltage and IP

The spectra from both open loop and sensorless controller are presented in Fig. 4 and 5 respectively. The comparison between current, voltage and combination current and voltage termed as instantaneous power (IP) in sensorless controller clearly indicate that the IP spectrum shows distinctive features. For the current spectrum, the gradual increased on amplitudes of slip-frequency of the spectra is clearly identified as the loads are increased. However, for voltage spectra, it is difficult to obtain the differences.

In general cases the spectra from figures 4 and 5 show that two (lower and upper) sidebands are asymmetric. The spectra show the more $2sf_s$ (twice slip-frequency) amplitude on the lower than upper sideband frequencies. This agree with [9] that there are new additional components occur under BRB condition. The two components are situated on the lower sideband frequency with different phase and the other one component lies in the upper sideband frequency.

However, on comparing the two modes, the sensorless mode gives a better view than to the volts/Hz mode.

4.2 Detection and Diagnosis Features

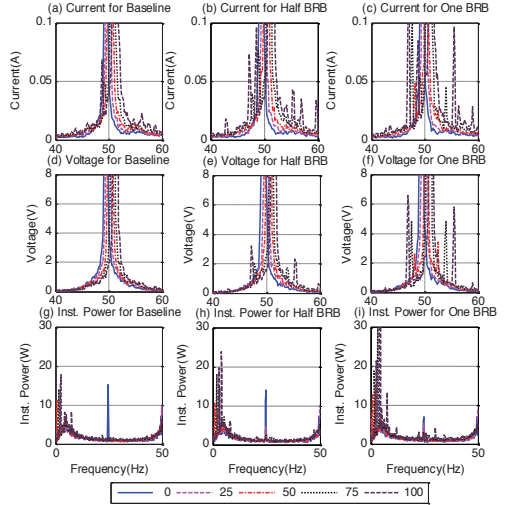


Figure 5. Spectra of current, voltage and IP for different loads under sensorless operation

Figure 6 shows the diagnostic result comparison for healthy, half broken bar and one broken bar on volts/Hz operating mode at variable loads from no load to full rated load. The rms values were extracted from the current, voltage and extraction from both current and voltage signals to calculate IP.

At high load shows the distinct values of current, voltage and IP from healthy and faulty machines. IP vs load graph has given the best distinct values for fewer loads applied to the IM. The fault prediction on this case is easier for the load above 25% load.

On the other hand, Figure 7 shows diagnosis results from the sensorless operation. In generally, it gives better results on both current and voltage vs load graph compared to the volts/Hz mode. However for IP vs load graph, sensorless operation even gives more distinctive results for healthy and faulty condition on wider range of loads.

The twice slip frequencies for each load shifts away from the fundamental frequency (f_s) as the load increases. This is means that the slip increases when the load increases.

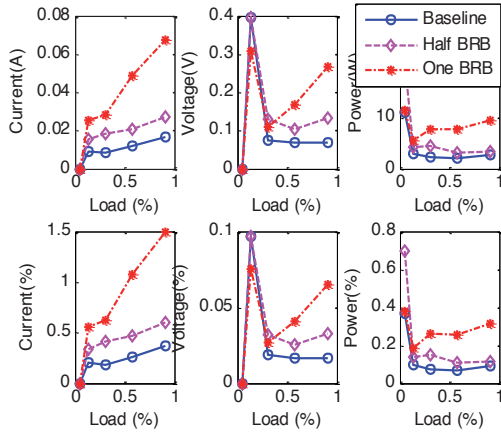


Figure 6. Diagnostic result comparisons from volts/Hz operation

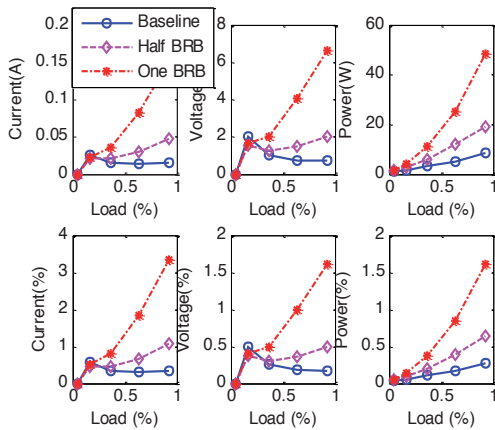


Figure 7. Diagnostic result comparisons from Sensorless operation

5. CONCLUSION

BRB faults can be detected on sensorless mode of operation the diagnosis of BRB based on VSD. The analysis of the VSD process has understood that the sensorless control mode adjusts both supply voltage and current in order to maintain the operating speed within high accuracy. It means that any asymmetric problems with the rotor such as BRB will cause changes to current, voltage and power signals. Therefore, all of these signals can be used to diagnose BRB.

Experimental studies show that the components at twice slip frequency can be extracted more reliably in the spectrum of the power signal which contains diagnostic information of both current and voltage signals. Consequently, it produces better results in separating different BRB cases under sensorless control mode, compared with that of using either current or voltage signal.

It is shown that current signal of the faulty (BRB) case exhibits three new additional components compared to the healthy one. Two of these are the lower sideband components at the same frequency, but with different phases, and the other is the upper sideband component with a phase different from the previous two. That is why the lower sideband frequency is higher in amplitude compared to the higher one.

In addition, sensorless results: IP gives the better results in differentiating BRB because it includes the information from both current and voltage signal both of which are adjusted by the drive and especially it is possible to distinguish the faults at very low load.

6. APPENDIX

MOTOR DATA:

Base Frequency	: 50 [Hz]
Motor Voltage	: 400 [V]
Motor Current	: 9.2 [A]
Nameplate RPM	: 1420 [RPM]
Motor Poles	: 4 poles
Power	: 4 [KW]
Motor Connection	: star
Power Factor	: 0.84
Stator Resistance	: 0.9904 [Ω]
Leakage Inductance	: 19.22 [mH]
Mutual Inductance	: 129.49 [mH]
Rotor Time Constant	: 114.81 [ms]

7. ACKNOWLEDGMENTS

Authors express our thanks to The Directorate General of Higher Education Ministry of Education and Culture Republic of Indonesia for providing the financial support.

8. REFERENCES

- [1]. Bellini, A., et al., *Advances in Diagnostic Techniques for Induction Machines*. Industrial Electronics, IEEE Transactions on, 2008. 55(12): p. 4109-4126.
- [2]. Mehriou, M.R., et al., *Rotor fault condition monitoring techniques for squirrel-cage induction machine—A review*. Mechanical Systems and Signal Processing, 2011. 25(8): p. 2827-2848.
- [3]. Trzynadlowski, A.M., M. Ghassemzadeh, and S.F. Legowski, *Diagnostics of mechanical abnormalities in induction motors using instantaneous electric power*. Energy Conversion, IEEE Transactions on, 1999. 14(4): p. 1417-1423.

- [4]. Tavner, P.J., *Review of condition monitoring of rotating electrical machines*. Electric Power Applications, IET, 2008. 2(4): p. 215-247.
- [5]. Xi, Z., *Sensorless Induction Motor Drive Using Indirect Vector Controller and Sliding-Mode Observer for Electric Vehicles*. Vehicular Technology, IEEE Transactions on, 2013. 62(7): p. 3010-3018.
- [6]. Won Seok, O., et al. *Self-tuning speed controller for induction motor drives*. in *Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM), 2012 International Symposium on*. 2012.
- [7]. Toliyat, H.A. and T.A. Lipo, *Transient analysis of cage induction machines under stator, rotor bar and end ring faults*. Energy Conversion, IEEE Transactions on, 1995. 10(2): p. 241-247.
- [8]. Thomson, W.T. and M. Fenger, *Current signature analysis to detect induction motor faults*. Industry Applications Magazine, IEEE, 2001. vol. 7(4): p. 26-34.
- [9]. F. Gu, et al., *A new method of accurate broken rotor bar diagnosis based on modulation signal bispectrum analysis of motor current signals*. Mech. Syst. Signal Process, 2014.

eMaintenance 2A

Failure Prediction of Tidal Turbines Gearboxes

Faris Elasha
Cranfield University
Cranfield, Bedford
f.elasha@cranfield.ac.uk

David Mba
Cranfield University
Cranfield, Bedford
d.mba@cranfield.ac.uk

Joao Amaral Teixeira
Cranfield University
Cranfield, Bedford
.a.amaral.teixeria@cranfield.ac.uk

ABSTRACT

In order to truly minimize the maintenance cost and prevent failures of tidal turbine gearboxes, there exists a fundamental need for a prognostic tool that can reliably estimate the current health and reasonably predict the future condition of the gearbox. The research presented is aimed at developing a prognostic tool to predict the remaining life of the gearbox during operation and utilise this tool for maintenance planning. A prognostic model for the remaining life prediction of a gearbox has been developed. This model utilises the data collected by a monitoring system to predict the future condition of the gearbox. The result showed that applying real load condition results in reduction of time to failure initiation compared to average condition.

Keywords

Prognostic, Failure Prediction, Gearboxes.

1. INTRODUCTION

Globally, renewable energy generation has significantly increased during the last decade with an increment of 33% in the UK as of 2011 [1]. Marine energy is one of the promising energy resources, especially in the UK; estimation shows a capability for producing 36 GW from tidal energy alone. By 2020, Europe plans to increase the installed capacity of tidal turbines to 1.95 GW and 50% of these will be installed in the UK [2].

Operation and maintenance (O&M) decisions for tidal turbines contribute significantly to the cost of tidal energy. Such decisions generally depend on many factors such as machine health, repair costs, weather conditions, etc... These factors make O&M decision a challenging problem. Therefore, there is an urgent need for reliable planning tools [3].

Availability of tidal turbines significantly affects their economic viability, a key aspect of tidal turbine availability is the need for efficient planning for maintenance resources. Condition Monitoring Systems (CMS) are the answer for maintenance management and increased reliability [4-6]. Such systems are commonly used in other industries. They continuously monitor turbine components and provide an optimum maintenance scheduling.

Generally prognostic approaches can be categorised into three, data driven, physics based and a hybrid approach. The majority of current research in gearboxes prognosis uses data driven methodology which is based on vibration and oil analysis [7-9] technologies. Typically the data is collected during operation, and then statistically treated to estimate the residual life (RUL).

However the time between the residual life (RUL) prediction and failure is relatively short [10], which ultimately leads to higher maintenance cost. Physical based models have been applied for prediction of life based on crack propagation theory, such models require significant information and are difficult to develop. Therefore physical models are not established in industry [11].

This paper introduces a new prognostic approach for predicting the remaining life centered on a methodology that combines data and physics based models. The prediction of remain life of the gears is based on the combination of fracture mechanics (Physics model) and operational data (data driven model). The aim of this research is to introduce this new methodology as a practical tool for gearboxes prognosis.

2. Gearbox prognostic Model

2.1 Concept

The prognostic model aims to evaluate gearbox remaining life. The first stage of the model is focused on predicting the residual time before failure initiation. The schematic of the prognostic model is shown in Figure 1. The prediction is determined on gear life estimates based on BS.ISO 6336 [12]. In order to perform this estimate, speed and torque measurements are required, as well as certain design parameters. The torque and speed data are processed to estimate the load spectrum, which is used to estimate the residual life.

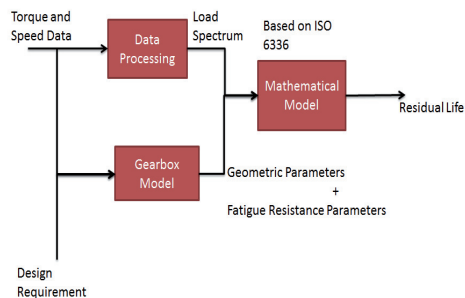


Figure 1. Gearbox prognostic model

The configuration considered in this study is used in wind turbines and known as GB-R100-A [13], and consist of a combination of two planetary stages followed by one parallel stage, the details of the design can be seen in table 1.

Table 1. Final geometric features of the 1st planetary stage gears

Gear	Sun	Planet(×4)	Internal Gear
Number of teeth	27	47	121
Module (mm)	8		
Pitch diameter (mm)	216	376	968
Helix angle (°)	0		
Facewidth (mm)	90		
Centre distance (mm)	296		

2.2 Load spectrum

Ideally the load spectrum of the tidal turbine transmission should be extracted from operational measurements of torque and speed. However, for purpose of this investigation data from numerical simulation of tidal flow was employed, in which speed measurements of the flow were used to estimate the rotor torque. The numerical simulation was based on Blade Element Momentum (BEM) theory [14]. Tidal follow and speed conditions were measured during one lunar cycle (28 days) [15].

As the load data correspond to the load on the turbine rotor, it was then possible to estimate the load in the sun and planet gears of the first stage. Equation (1) was used for this purpose, in which w and T denote revolution and torque respectively. The process of building load spectrum starts by estimating the bin size of the spectrum, and then relative frequencies of each interval are estimated based on duration spent on each loading interval. The duration of each interval is used to calculate the number of cycles according to the ISO 6336.6 recommendations as shown in equation 2 [16].

$$P_t = \omega_{rotor} \times T_{rotor} = \omega_{sun} \times T_{sun} \quad (1)$$

$$N_t = \omega_t \times 60 \times \frac{T_{sample}}{L_{sample}} \times F_{sample} \times L_{operation} \quad (2)$$

Where, N_t the number of cycles of one tooth of the gear will experience over duration.

ω_t The rotating speed of the gear during the corresponding load (rpm).

T_{sample} The time duration spent over certain load bin (interval).

L_{sample} The total time length of the data (s).

F_{sample} The fraction of time corresponding to the load under consideration.

$L_{operation}$ The design life of the turbine (h).

Most methods to estimate the bin size are statistical approaches. To obtain the best representative spectrum, the following approaches were considered:

- I. Application of the Scott Method [17] which states the best bin size depend on the data distribution. This approach provides an efficient estimation based on probability density distribution, the best bin size estimated by:

$$W = 3.49\sigma N^{-0.333} \quad (3)$$

Where, W is bin width, σ Standard deviation of distribution, N number of samples.

- II. Another more robust method for data mining is the Freedman and Diaconis method [18], which is considered as an improvement to Scott method, in which the standard deviation was replaced by interquartile distribution of data. The optimum bin size is determined by:

$$W = 2(IQR) N^{-0.333} \quad (4)$$

Where, IQR is the interquartile which depends on the probability distribution of the data.

- III. ISO 6336 recommends using of large bin sizes for low load ranges and small bin size for larger loads. However, no mathematical formulation for such estimates have been proposed by the ISO standard.

The data collected was classified into two, ebb and flood. Each class was assumed to represent 50% of the lifetime and each class contained 7 loading groups which represent the loading case for the whole life. As this data was collected during one lunar cycle (28 days), the probability density function was used as the basis for predicting the load experienced by gearbox throughout its life. The probability density of tidal velocity used for load estimation is shown in Figure 2, an example of a load cycle presented in Figure 3.

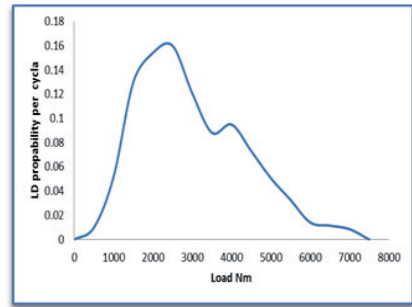


Figure 2. Probability density function for loading condition

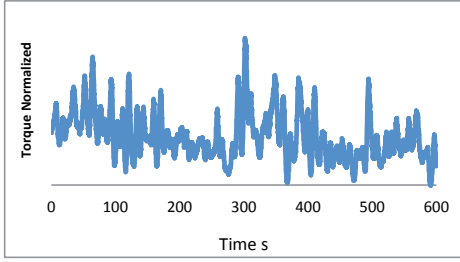


Figure 3. Example of load groups

3. Life Estimation

Fatigue resistance factors are required for life estimation; these factors are calculated using the ISO standard based on gear geometric and material specifications. These factors were introduced to take into account the influence of many characteristics of the gears such as the elasticity of the material, the helix angle of the teeth and the number of cycles in the design life. These factors were categorized into three, general influence factors (K), pitting resistance factors (Z) and bending resistance factors (Y). The general influence factors are used in both pitting and tooth bending resistance calculations. General influence factors include application factor KA, which accounts the effect of variable load, dynamic factor Kv, which makes allowance for the effects of gear tooth quality level and modifications as related to speed and load. Moreover, load factors ($K_{H\beta}$ and $K_{H\alpha}$) were considered to take into account the influence of load distribution on both normal and transverse directions [19]. The pitting resistance factors include geometry factors, which accounts for the influence of geometry characteristics to contact fatigue such as zone factor Z_H , helix angle factor Z_β etc. In addition, pitting fatigue resistance factors account for the effect of material and oil film, estimation methodology of pitting resistance factors is detailed in ISO 6336.2 standard [20]. The bending fatigue resistance factors determine the effect of geometry and surface condition on gear root bending fatigue, ISO 6336.3 is used to calculate these factors [12]. A numerical tool was used to extract this feature based on Method C of ISO 6336. The result of these factors are summarised in table 2.

Table 2. Estimated Application factor

Ebb only loads condition		
	Sun Gear pitting	Sun Gear Bending
Actual load	1.93	1.86
Average	0.844	0.815
Flood only loading condition		
Actual load	2.54	2.43
Average load	0.995	0.954
Flood and Ebb loading condition		
Actual load	2.26	2.17
Average load	0.934	0.894

3.1 Application factor

The application factor is used to quantify the effect of varying load on the fatigue life of the gears. This factor compares the varying load to steady loading; as yet, this factor has not been established for tidal turbines due to the lack of experience and data in this field. Data simulated using Blade Element Momentum theory [14] was used to establish the torque data for range of flow conditions [15]. The procedure used to estimate the application factor was established in ISO 6336.6 Appendix A [16]. The results of application factor estimation are summarized in Table 2.

Table 3. Fatigue resistance factors calculated using a numerical tool (KISSOFT™)

Parameter	Sun gear	Planets gears
Dynamic load K_V	1.001	
Transverse load factor (contact stress) $K_{H\beta}$	1.063	
Face load factor (root stress)s $K_{F\beta}$	1.049	
Face load factor (contact stress) $K_{H\alpha} / K_{F\alpha}$	1	
Zone factor Z_H	2.495	
Single pair tooth contact factors $Z_{B/D}$	1.0	1
Elasticity factor Z_E	189.812	
Contact ratio factor Z_ϵ	0.878	
Helix angle factor (contact) Z_β	1	
Life factor (contact) Z_{NT}	0.95	0.972
Lubricant factor (contact) Z_L	1.047	
Velocity factor Z_V	0.942	
Roughness factor Z_R	0.99	
Work Hardening factor Z_W	1	
Size factor Z_X	1	
Tooth form factor Y_F	1.39	1.290
Stress correction factor Y_S	1.92	2.045
Stress correction factor Y_{ST}	2	
Helix angle factor (tooth root) Y_β	1	
Rim thickness factor Y_B	1	1
Deep tooth factor Y_{DT}	1	
Life factor (tooth root) Y_{NT}	0.91	0.928
Test relative notch sensitivity factor $Y_{\delta_{relT}}$	0.99	0.996
Relative surface factor Y_{RrelT}	1.04	1.047
Size factor (tooth root) Y_X	0.97	0.97
Mean stress influences factor Y_M	1	1
Safety factors in pitting	1.25	1.313
Safety factors in tooth bending	2.56	2.652

In order to estimate the gear life, bending and pitting stress spectra are required. This bending stress spectrum is calculated by [12]:

$$\sigma_{fi} = 2000 \frac{T_i}{d_{ref} m_n b} Y_s Y_F Y_\beta K_{vi} K_{Bi} K_\alpha \quad (1)$$

The pitting stress spectrum is estimated by [20]:

$$\sigma_{Hi} = Z_H Z_E Z_\epsilon Z_{\beta D} \sqrt{2000 \frac{T_i}{d_{ref}^2 b} \frac{u+1}{u} K_{v_i} K_{H\beta} K_{H\alpha i}} \quad (2)$$

Then the nominal stress spectra for pitting σ_H and bendig σ_F estimated by:

$$\sigma_H = Z_B \sigma_{Hi} \sqrt{K_A K_V K_{H\beta} K_{H\alpha}} \quad (3)$$

$$\sigma_F = \sigma_{Fi} K_A K_V K_{F\beta} K_{F\alpha} \quad (4)$$

These stress spectra are used to estimate the life factors for pitting Z_{nt} and bending Y_{nt}

$$Z_{nt} = \frac{\sigma_H}{\sigma_{HP}} \quad (5)$$

$$Y_{nt} = \frac{\sigma_{fi}}{\sigma_{fp}} \quad (6)$$

Which in turn is used to estimate the corresponding number of cycles to failure for each load bin using graphical information in ISO 6336-2:2006, figure 6, and ISO 6336-3:2006, figure 9. Then the damage due to fatigue is calculated for each cycle using Miner's rule [21]:

$$D = \sum_{i=0}^n \frac{N_i}{N_t} \quad (7)$$

In which N_t is the number of cycles of one tooth of the gear experienced, and N_i is the total number of cycles to cause damage under corresponding loading condition which is estimated based on material SN curve (ISO 6336-2:2006, Figure 6, and ISO 6336-3:2006, Figure 9) and stress experienced.

4. Model result and sensitivity analysis for life analysis

Simulated load data was used for the life calculation is shown in Figure 4; the calculations started by estimating load spectra, and then converting it to a stress spectra on the gears as shown in Figure 6. The stress spectra was used to obtain the Miner sum using geometric features, fatigue resistance and equations described previously. Different bin sizes were used to perform the sensitivity study. This means that for each bin size, different application factors and spectra were estimated.

Observations of Figure 7 and Figure 8 show that the life estimation based on Miner sum varies depending on bin size, the results converge to constant value as the number of bins is increased. Using the Freedman and Diaconis, Scott and ISO methods. optimum bin size (introduced in section 3) results in similar life predictions to smaller bin size.

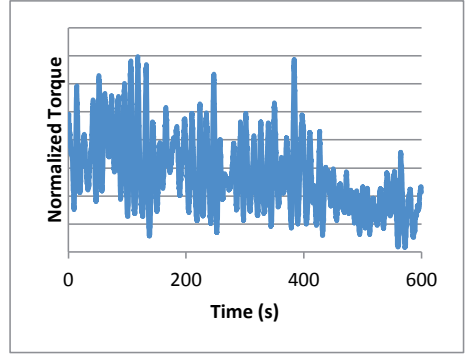


Figure 4. Example Torque Data

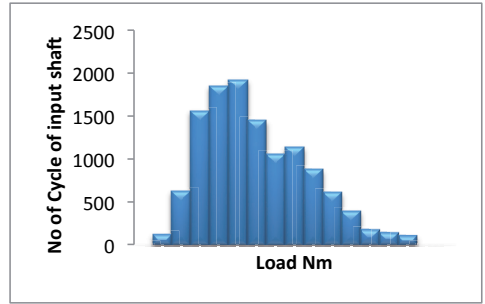


Figure 5. Load stress spectra

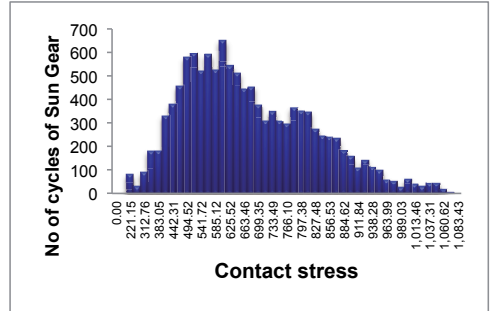


Figure 6. contact Stress Spectra of sun gear

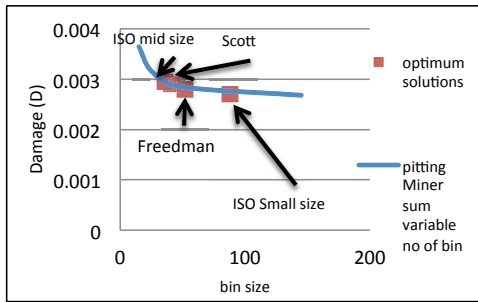


Figure 7. Estimates of damage for pitting stress (one data set)

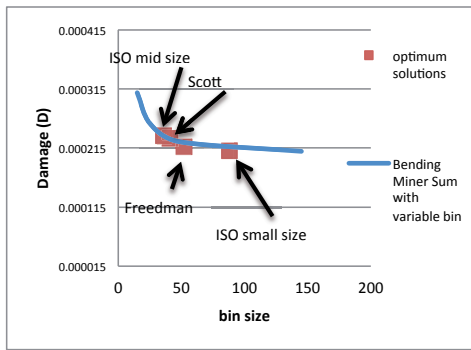


Figure 8. Estimates of damage for bending stress one data set)

In order to validate of the prognostic model, the first stage gears life was estimated for both rated and simulated load conditions, see table 4. Observations of this table show the actual life of the first stage gears is less compared to design life due to the varying load effect, therefore it is recommended to design the gearbox using realistic loading condition. From Prognostic point of view using the actual load data allows for the continues evaluation of the gears fatigue strength. The gears will be considered with initial damaged when the miner sum D reaches 1.

Table 4. Life Prediction

Method	Designed life Average load	Freedman	Scott
Predicted life	1.92×10^8 cycles	8.84×10^7 cycles	8.68×10^7 cycles

5. Validation

The validation of this concept is based on pitting tests performed by Khan et. Al. [22], in which two pitting tests were performed on two identical pairs of case-hardened low carbon steel gears, the gears tested under two loading conditions and the useful life was estimated based on ISO 6336-2 guidelines described previously.

The details of gears used and load condition are summarized in table 5.

Table 5. Gear Parameter according to design calculation (ISO6336.2) [22]

Gear	Specification
Helix angle	17.75
Center to distance	113
No of teeth	35
Reference diameter	110.2
Load condition 1	6500 N
Load Condition 2	4347 N
Speed	1000 RPM
Life estimated in load condition 1	5.985×10^5
Life estimated in load condition 1	9.67×10^5

According to the test result the pitting was detected during the test when the number of cycles reached 5.4×10^5 cycles during the first load condition and 1.08×10^6 cycles during the second load condition. Applying of Miner sum using equation 3, the damage level was estimated at 0.92 and 1.034 for test 1 and test 2 respectively. Visible pitting after gear testing is shown in Figure 9 and Figure 10.

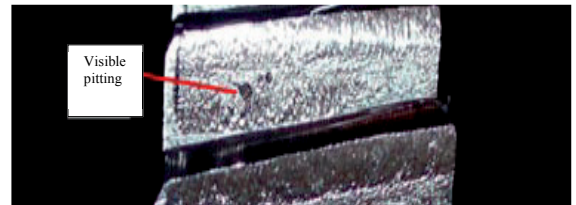


Figure 9. Visible pitting after test 1 [22]



Figure 10. Visible damage after test 2 [22]

Observations of Figure 9 and Figure 10 show small pits were noticeable at the Miner sum (damage level) of 1. This result proves the validity of mathematical model of BS-ISO 6336-2 in life prediction.

Further research is required to validate this model for conditions of time varying load. In addition, the size of pits detected in the

validation experiment shows that the gears can be usefully operated for a significant duration after damage initiation.

6. Conclusion

A prognostic model based on the loading condition has been developed to predict the residual life of the gearbox during turbine operation. The model used load and speed measurements to estimate the current health and predict the remaining life. The accuracy of the prognostic model was investigated by a sensitivity study, the effect of load data processing was studied, and the result shows that applying a different bin width of load data can significantly affect the accuracy of the prediction, however application of statistical methods to select the optimum bin size result in accuracy improvement. The model was validated using constant load pitting test data and accurate prediction of life was proved. The results raise the need to develop this model further to include pitting progress prediction to assist in deciding the time for repair.

7. References

- [1] Digest of UK energy Statistics 2012, (2012), 12/089, Department of Energy and Climate Change, UK.
- [2] Ben Elghali, S. E., Benbouzid, M. E. H. and Charpentier, J. F. (2007), "Marine Tidal Current Electric Power Generation Technology: State of the Art and Current Status", Electric Machines & Drives Conference, 2007. IEMDC '07. IEEE International, Vol. 2, pp. 1407.
- [3] Wang, P., Tamilselvan, P., Twomey, J. and Youn, B. (2013), "Prognosis-informed wind farm operation and maintenance for concurrent economic and environmental benefits", International Journal of Precision Engineering and Manufacturing, vol. 14, no. 6, pp. 1049-1056.
- [4] Kacprzynski, G. J., Roemer, M. J. and Hess, A. J. (2002), "Health management system design: Development, simulation and cost/benefit optimization", Aerospace Conference Proceedings, 2002. IEEE, Vol. 6, pp. 6-3065.
- [5] Rudov-Clark, S., Stecki, J. and Stecki, C. (2011), "Application of advanced failure analysis results for reliability and availability estimations", Aerospace Conference, 2011 IEEE, pp. 1.
- [6] Tavner, P. J., Xiang, J. and Spinato, F. (2007), "Reliability analysis for wind turbines", Wind Energy, vol. 10, no. 1, pp. 1-18.
- [7] Mba, D. (2005), "Prognostic opportunities offered by acoustic emission for monitoring bearings and gearboxes", Twelfth international congress on sound and vibration, .
- [8] Dupuis, R. (2010), "Application of oil debris monitoring for wind turbine gearbox prognostics and health management", Annual Conference of the prognostics and health management society, pp. 10.
- [9] Kitaljevich, D., Dupuis, R. and Lu, M. (2009), "Oil Debris Condition Monitoring for Wind Turbine Gearboxes", Beijing Wind Power, 21-23 October, Beijing, China, .
- [10] Reuben, L. C. K. and Mba, D. (2014), "Diagnostics and prognostics using switching Kalman filters", Structural Health Monitoring, vol. 13, no. 3, pp. 296-306.
- [11] Aslantaş, K. and Taşgetiren, S. (2004), "A study of spur gear pitting formation and life prediction", Wear, vol. 257, no. 11, pp. 1167-1175.
- [12] British Standards, (2006), Calculation of load capacity of spur and helical gears - Part 3: Calculation of tooth bending strength (BS ISO 6336-3), UK.
- [13] Smolders, K., Long, L., Feng, Y. and Tavner, P. J. (2010), "Reliability Analysis and Prediction of Wind Turbine Gearboxes", European Wind Energy Conference, 20-23 April 2010, Warsaw, Poland, UK, .
- [14] Masters, I., Chapman, J., Orme, J. and Willis, M. (2011), "A robust blade element momentum theory model for tidal stream turbines including tip and hub loss corrections", Proceedings of IMarEST-Part A-Journal of Marine Engineering and Technology, vol. 10, no. 1, pp. 25-35.
- [15] Adrian, L. (2012), Gearbox design for tidals turbines (unpublished MSc thesis), Cranfield University, UK.
- [16] British Standards, (2006), Calculation of load capacity of spur and helical gears - Part 6: Calculation of service life under variable load BS ISO 6336-6, BSI, UK.
- [17] Scott, D. W. (1979), "On optimal and data-based histograms", Biometrika, vol. 66, no. 3, pp. 605-610.
- [18] Izenman, A. J. (1991), "Review Papers: Recent Developments in Nonparametric Density Estimation", Journal of the American Statistical Association, vol. 86, no. 413, pp. 205-224.
- [19] British Standards, (2006), Calculation of load capacity of spur and helical gears - Part 1: Basic principle, introduction and general infulnace factors BS ISO 6336-1, BSI, UK.
- [20] British Standards, (2006), Calculation of load capacity of spur and helical gears - Part 2: Calculation of tooth contact strength (BS ISO 6336-2), BSI, UK.
- [21] P., R. (2012), Dudley's Handbook of Practical Gear Design and Manufacture, second ed, CRC press, USA.
- [22] Khan, M. A., Cooper, D. and Starr, A. (2009), "BS ISO Helical Gear Fatigue Life Estimation and Wear Quantitative Feature Analysis", Strain, vol. 45, no. 4, pp. 358-363.

A web service-based toolbox for machine diagnostics based on statistical analysis

Luca Fumagalli
Department of Management,
Economics and Industrial
Engineering,
Politecnico di Milano,
P.zza Leonardo Da Vinci 32,
20133 Milano, Italy
luca1.fumagalli@polimi.it

Simone Pala
Department of Management,
Economics and Industrial
Engineering,
Politecnico di Milano,
P.zza Leonardo Da Vinci 32,
20133 Milano, Italy
simone.pala@polimi.it

Marco Macchi
Department of Management,
Economics and Industrial
Engineering,
Politecnico di Milano,
P.zza Leonardo Da Vinci 32,
20133 Milano, Italy
marco.macchi@polimi.it

ABSTRACT

The purpose of this research is to present a solution as a case study to discuss the design choices of a flexible platform based on the combined use of ICT developments – as the web services – and statistical analysis – as a Black-box approach –, in order to obtain a tool-box featuring high usability. The innovation proposed by the paper concerns the possibility to make the analysis in a way that allows interoperability with other IT systems of the company or the possible integration with other maintenance modules. On the other hand, diagnostic techniques adopted in the research are consolidated statistical approaches. Overall, web services are a Key Enabling Technology for agile diagnostics tasks, and the present work exploits the advantages offered by this technological trend.

Keywords

e-Maintenance, Black-box approach, web-service, diagnostics.

1. INTRODUCTION

During the last years, e-Maintenance solutions have undergone a remarkable evolution: on one hand, the growing level of efficiency and availability of assets has become a crucial need for industry to reduce costs, leading to an increasing need for new solutions for maintenance; on the other hand, ICTs (Information and Communication Technologies) have increased the possibility of providing modular and flexible solutions.

One of the most preferred e-Maintenance applications is the Condition Based Maintenance (shortly referred to as e-CBM): it is commonly considered as technically feasible and financially viable, and it offers the expectation of high profits, by reducing significantly a plant downtime (Djurdjanovic et al 2003). Campos (2008) makes a review of the literature on the application of e-CBM, using the OSA-CBM layers (Open System Architecture Condition-Based Maintenance) for the analysis of the reviewed work.

Diagnostics is a key activity in CBM and it focuses on detection, isolation and identification of faults either once occurred or before occurring (Jardine et al., 2006). It is a procedure mapping the information obtained in the measurement space and/or features in the feature space to machine faults in the fault space. Following this principle, and matching it with the MIMOSA framework (MIMOSA, 2010), it is possible to identify the functionalities of a generic CBM tool: the State Detection, the Health Assessment (with fault isolation and identification), the Prognostics (extending

the diagnosis based on a remaining useful life estimate) and, last but not least, the Advisory Generation.

Different key enablers have emerged in the last decades, boosting e-Maintenance thanks to the new technological solutions and components available from the ICTs development trend. Indeed, service-based solutions seem promising in terms of flexibility and interoperability, to finally become a Key Enabling Technology to support agile diagnostics.

2. ROLE OF WEB-SERVICES FOR MAINTENANCE

The development of information system for data management within maintenance has been emphasized by the adoption of web-services, allowing to spread out e-Maintenance strategies.

In fact, web-services offer different advantages. Firstly, they allow to make available on the network any kind of information, also the one related to maintenance. In particular, the web service technology permits the interoperability of heterogeneous devices within the same network, and it also allows to access information from a normal web-browser. Moreover, web services may work as a base to build Software as a Service (SaaS) (NIST, 2011). Hence, the availability of web-based CMMS allows to embed information coming from the field as exposed by web-services within the CMMS pages.

Thus, it can be foreseen a future of maintenance information system oriented to embed data and information coming from the field. To this end, the following brief excursus on maintenance information system evolution may be beneficial, to understand the novelty.

In the 70's the first computerized system started supporting preventive maintenance. With the advent of microcomputers, new functionalities were developed in the 80's, such as the management of archives, budgets and stocks, which reduced plant downtime and faults. All these improvements paved the way for the advent of the first CMMS at the end of the 80's. The exchange and management of information and the creation of databases characterized this phase. In the late 90's, after the development of MRP (Material Resource Planning) systems, companies began to pay more and more attention to integrated solutions such as ERP (Enterprise Resource Planning). From this moment on, the information systems have been integrated with all the other managerial sectors of a company; therefore, collaboration and exchange of information became leading concepts.

As it has been said, in the last years, information technologies have developed considerably and, since 2000, e-Maintenance logics, together with the implementation of web-service solutions in maintenance management, have increased this trend. Definitely, it is possible to state that the current and future development of maintenance management systems aims at obtaining systems that can be integrated more and more with other platforms, including CBM platforms and tools.

3. THE PROPOSED TOOLBOX

Different approaches are possible to deploy CBM functionalities. A first classification concern two alternative type of approach: “White-box” and “Black-box” (Ierace et al., 2009). The former concerns the physical modelling of an equipment behavior in order to have a basis for the diagnostic analysis. The latter concerns mathematical modeling based on statistical data, hence without the support of physical modeling. Statistical approaches (e.g. Jain and Duin, 2000) are indeed considered representative of a Black-box approach.

The two approaches have both pros and cons; nevertheless, Black-box approaches are suitable in all those situations when physical system models are not available or, at least, they are too difficult to be customized to the specific equipment and operating context. In this sense, a Black-box approach may be claimed to be flexible enough to be exploitable in a larger number of cases. To this end, a research on Black-box approaches is deemed by the authors to be highly interesting for wide application range in large industrial contexts.

The research grounds on ProDEST (Prognostics and Diagnostics based on Electric Signature Technique) introduced by Fumagalli et al., 2011, a toolbox currently under development by the LGM (Laboratorio di Gestione della Manutenzione) of Politecnico di Milano, DIG (Department of Management, Economics and Industrial Engineering), basically consisting in a series of independent modules offering a set of features capable of supporting CBM.

This paper presents the modules that support State Detection and Health Assessment features. In order to provide advancements with respect to the state of the art, the tool is: (i) based on statistical techniques, such as different types of Control Charts and Principal Component Analysis; (ii) structured with the paradigm of Web service to allow modularity and interoperability with other ICT systems. More specifically with concern to data collected from the field, the toolbox is mainly devoted (but not limited) to process electrical signals, acquired by transducers installed on the electro-mechanical systems, for machine fault diagnosis.

Figure 1 presents the MIMOSA OSA-CBM framework used as a reference to build the CBM tool, while Figure 2 provides an overview on the functional architecture of the proposed toolbox and its correspondence with MIMOSA OSA-CBM. The toolbox presented is focused on State Detection (SD) module that is discussed in the current paragraph. The remaining necessary blocks (DA, DM, HMI & external systems) that constitute the toolbox, and their interconnections, are presented in the next paragraph.

More specifically, in this section some statistical tools that can compare different states of the equipment are considered. In this way, it is possible to identify faults in a rigorous manner. However, to this end, it is necessary to have a high amount of data

coming from tests made on the machines. Companies collect data very often, so it is possible to elaborate them efficiently in order to obtain information about the system under examination.

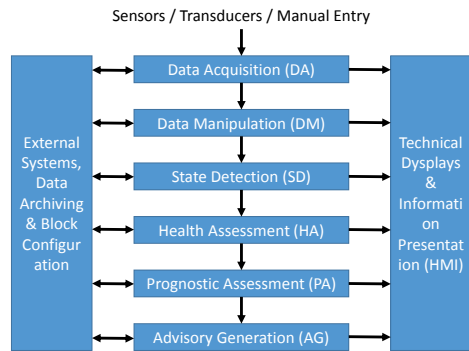


Figure 1. MIMOSA OSA-CBM: ISO-13374 data processing and information flows

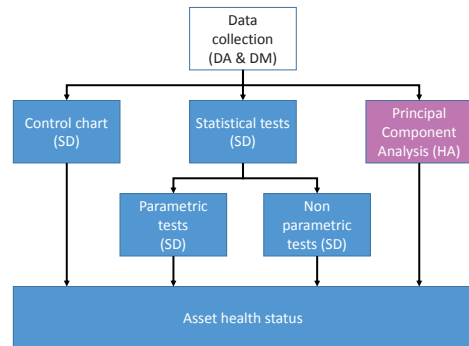


Figure 2. Functional architecture of the proposed toolbox and correspondence with MIMOSA OSA-CBM

A first analysis is possible through histograms (bar graphs) and by calculating the main statistical parameters, such as the mean of the statistics, the standard deviation (to evaluate data dispersion), skewness (to measure the deviation from distribution from symmetry), kurtosis (as an index of “peakedness” of distribution) and other synthetic parameters. This kind of indexes can be regarded as excellent means to approximately identify the problem. Yet, the tools that will be discussed in the following paragraphs should be employed to understand better the situation under consideration (Ageno, 2013), thus helping SD.

3.1 Control charts

A control chart is a graph that consists of upper and lower control limits, drawn as separate lines: a central line is drawn at the value of the mean of the statistic of process outputs and points represent each measured value (Galletto, 2000).

The fluctuation of the points within control limits is due to the intrinsic variability of the process itself. On the contrary, the

points that fall outside the control limits may be signalling the presence of a special cause that requires immediate investigation and elimination (Galletto, 2000).

Control charts fall in two categories: variable data control charts and attribute data control charts. The former are used when one has data that can be measured on a continuous scale, while the latter are used with data that can be counted, such as defective ones, possessing or not possessing specific features.

From a CBM perspective based on measure on a continuous scale, we will focus on variable data control charts.

Ultimately, they can be considered as a monitoring technique that allows including in a graph all the information concerning a process.

As described in (Roberts, 1966), the first step to build a control chart is the examination of information, grouped into different observation X_i . One has to bear in mind that at the instant i it is possible either to collect single values or to calculate the mean of N measurements. Then, it is necessary to decide which kind of chart is to be used, according to the calculation method chosen to build the graph. Finally, the control limits and the relative control rules are established.

In the tool, three kinds of control charts have been implemented:

- Shewart control chart
- Uniformly Weighted Moving Average chart, UWMA chart
- Exponentially-Weighted Moving Average chart, EWMA

3.1.1 Shewart control chart

It is the typical control chart, which was introduced by Shewart in 1924 (Roberts, 1966). In this chart, the observations X_i are the mean value of N measurements taken at time i . Normally, they are supposed to be distributed with mean value μ and standard deviation σ_x . However, it is possible to build control charts directly from the collected data, that is to say $N = 1$ (namely one measure) for each time i .

This control chart is used to monitor and detect changes in the mean value of a single feature. The points on the chart represent the mean value of each sub-group (or the sample itself if $N = 1$). According to this chart the process is “out of control” when X_i falls outside the limits, in other words from a maintenance point of view, one can state that the monitored machine needs an action to solve the issue or further analysis to isolate the cause. It is important to underline that only the i -th observation is involved in decisional process. Shewart suggests to take as control limit $\mu \pm 3\sigma_x$, which is a typical value supported by the empirical nature of quality control.

The procedures introduced by Shewart is quite simple, but it proves to be particularly slow in suggesting a corrective action when the mean value of process μ is close to the reference value μ_0 (Roberts, 1966). In order to improve the individuation of slight mean-shift (difference between the measured mean of the process and the reference one), there are different ways to understand better the control chart.

The Shewart chart is the most diffused one and is efficient in identifying rapid jump of the monitored signal. On the other hand, small deviations of the signal are not identified.

One weakness point of this technique is that generally one chart for each variable is necessary. Moreover, many authors studied another important drawback, that is the effects of the lack of normality within control chart.

The common limits used in the case of normality of data seem, in fact, robust only if the data are not “extremely” not normal. In many cases samples of 4 or 5 individuals (data) are enough, unless high not normality is present. Nevertheless, normality of data is not often guaranteed, due to intrinsic distribution of data. To this end, UWMA and EWMA charts are more robust to the lack of normality and should be preferred.

3.1.2 UWMA and EWMA control chart

They are control charts grounding on different types of mobile average. UWMA stands for uniform weighted mobile average, while EWMA stands for exponential weighted mobile average. They express their maximum potential in identifying small changes in the signal and in working when normality of data is doubtful (Roberts, 1959).

The basic concept involved by these charts is the exploitation of the capability of treat data to minimize the probability of false alarms. In Ageno (2013) further details on how to build UWMA control chart are provided.

3.2 Statistical tests

When dealing with an industrial process, there are many disturbing variables. If they depend on each other, the resulting distribution will be approximately normal [9]. Often, the lack of knowledge of the process and the difficulty in managing some disturbing variables lead the analyst to collect normal control variables to simplify the examination of the process. This is an easier method, yet it proves to be a less effective one.

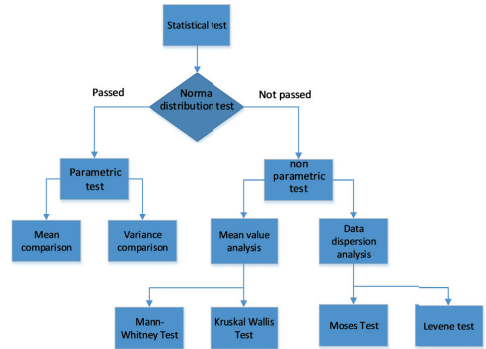


Figure 3. statistical test implemented in the toolbox

Therefore, the first thing to do is to verify the kind of distribution under consideration: only in the case of normal distributions it is possible to employ parametric statistical tests that give results that are easier to understand. Otherwise, it is preferable to apply non-parametric statistics, since they give results that are reliable regardless the kind of distribution.

Figure 3 provides a quick overview on the different statistical test (Mann-Whitney test, Kruskal-Wallis test, Moses test, Levene test and parametrical tests) implemented in the State Detection module

of the toolbox, providing a quick reference on their application cases.

3.3 Principal Component Analysis

The Principal Component Analysis, also called PCA, is a technique of multivariate not-parametric analysis that is used to simplify the original data with the purpose of reducing the number of variables to be considered. It may be used as tool for the classification of variables, individuation of outliers and the indications of anomalies in the data structure (Maere et al., 2012).

PCA applies a linear transformation to the original data matrix, in order to find a new set of variables (principal components) that express the maximum variability of data, with a reduced dimension than the original data and allowing to lose the minor quantity of information.

Then, a set of data may be confused, namely may be redundant (i.e. with correlated variables) and with a lot of noise (Maere et al., 2012). PCA serves, indeed, to eliminate such redundancy of the information, namely self-correlation.

The new reference system is subsequently constituted by the same variables, ranked with decreasing variance (Baydar et al., 2001). The first axis, thus, represents the variable with higher variance, the second axis the one with the second higher variance and so on.

The advantage in using a new reference system is to allow the possibility to highlight more easily the structure of the data (see Figure 4)

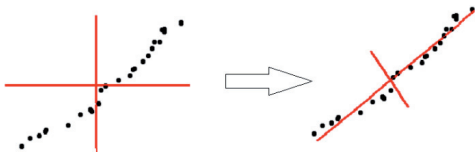


Figure 4. PCA geometric representation

It is worth noticing that, by applying PCA, it is possible to get a good trade-off between simplification of the analysis and information loss, and the right number of principal components to be considered is crucial to this end.

PCA has been used within the proposed architecture for an elaboration of the information to identify a proper reference space to allow fault identification within such space and thus proper Health Assessment.

4. THE ARCHITECTURE OF THE IMPLEMENTED CASE

Each part of the proposed tool has been coded in a modular way, allowing in this way to build a modular toolbox and, in particular, making easier the implementation of whole system by using web services. In fact, each module can be connected to another one simply following the “traditional way”, and so coding together the functions. Otherwise, it is possible to connect the different modules by using web services. The latter way allows also to easily “spread” the code and functions among different devices.

The above-described system has been implemented in a real case and the components are shown in Figure 5.

A case study has been carried out to demonstrate the flexible use of the developed toolbox in an industrial context.

A milling machine is selected as test case. The milling machine, as usual for this kind of system, may operate according to very different working cycles. For this reason, data acquired from a standard test cycle (i.e. simulating working condition but without carrying out any work on real pieces) were considered to define repeatable operating conditions. Validation phase provided positive results, thus confirming the interesting capability of the proposed toolbox.

The milling machine is equipped with its own sensors, so data are collected by the control system itself and, in this particular case, it is possible to transfer information to operators’ computer simply by performing a download-like action. In different scenario cases the data collection may be performed by data acquisition devices properly equipped with the right transducers and the needed software. In this specific case the milling machine control system performs DA and DM, then data are moved to the operator’s computer with a different DA function.

SD function is executed on the operator’s computer; it can be launched by an operator’s command and the results are shown on the screen or, since the function is built according web-service specification, it allows other devices on the network to request the execution of DA and SD, and to get the respective output.

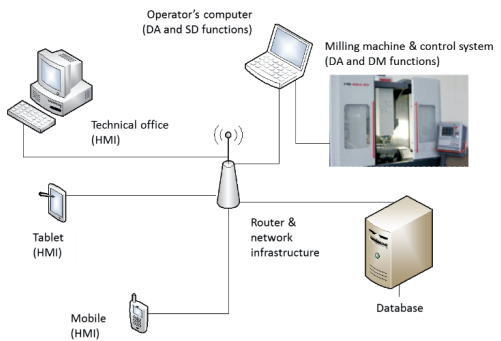


Figure 5. Hardware architecture of the implemented system

In particular, a data computation request on a web-service can be done by a device connected to the network, but also by a human through a common browser. The output provided is typically an XML string (<http://www.w3.org/XML/>) that is not easy to read by a human being; to overcome this limitation the output from a web service can be formatted into an human-readable page by using HTML (<http://www.w3.org/html/>), JavaScript code but also more advanced web interface language provided, for instance, by the software used to program the tool (i.e. LabView software by NI).

The so-realized architecture allows an operator to get information on the milling machine health status by using the toolbox on his computer, so standing next to the machine and accessing a “local” HMI; but it also allows to access the software interface from another PC (via browser), to execute data elaboration and see the results on a “remote” HMI. Technically speaking, we can define it as “Software as a Service” (SaaS).

Other devices in the network, such as the database – shown again in Figure 5 –, can connect to the operator’s computer to retrieve

data and store them for a future use. Also the communication among these two devices is regulated by web-service interfaces.

5. CONCLUSION

The value of this research is based on the architecture proposed, more than on the statistical functionalities that have been instead derived from consolidated statistical approaches. Indeed, the paper showed a possible implementation of an e-Maintenance toolbox based on web service technology focusing, in particular, on State Detection function.

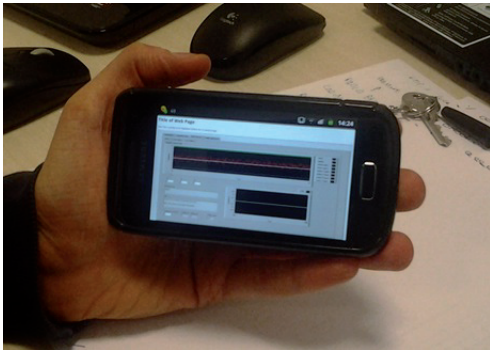


Figure 6. remote access to the tool via mobile

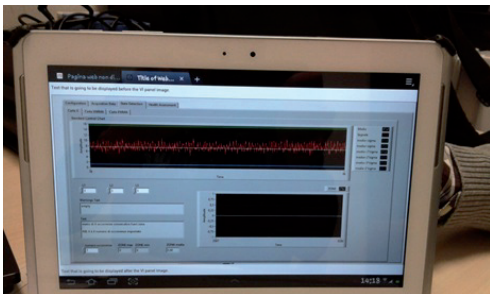


Figure 7. remote access to the tool via tablet

The aim is to show how it is possible to build the system and easily use it thanks to the adoption of web-services. From this point of view it is worth mentioning that it is possible to connect also mobile devices to the proposed architecture, see Figure 6 and 7. This extended capability makes the toolbox fully compliant with the recent scientific vision on e-Maintenance platforms.

The authors envision that future researches in this scope will address the deployment of the CBM functionalities by web-services, thus allowing to prepare different building blocks for a modular configuration of modern e-maintenance platforms.

6. REFERENCES

- [1] Baydar, N., Chen, Q., Ball, A., Kruger, U. (2001). "Detection of incipient tooth defect". Mechanical Systems and Signal Processing, pp. 303-321.
- [2] Djurdjanovic, D., Lee, J. and Ni J. (2003) "Watchdog Agent— an infotonics-based prognostics approach for product performance degradation assessment and prediction". Advanced Engineering Informatics 17.3, pp. 109-125.
- [3] Campos, J., Prakash, O., Jantunen, E. (2008). "A mobile maintenance decision support system. Maintenance and asset management". Vol 23 no 2, March/April 2008.
- [4] Fumagalli, L., Ierace, S., Dovere, E., Macchi, M., Cavalieri, S., Garetti, M. (2011). "Agile diagnostic tool based on electrical signature analysis". In Proceedings of IFAC 18th World Congress, August 28th – September 2nd 2011, Milan, Italy.
- [5] Galletto (2000), "QUALITA' Alcuni Metodi Strategici da Manager", Torino: C.L.U.T..
- [6] Ierace, S., Garetti, M. and Cristaldi, L. (2009). "Electric Signature Analysis as a cheap diagnostic and prognostic tool". Proceedings of WCEAM2009 (World Congress on Engineering Asset Management), Athens, Greece, 28 - 30 September 2009.
- [7] Jain Anil, K., Duin, R. P. W. and Jianchang, M. (2000). "Statistical Pattern Recognition: A Review". IEEE Transactions on pattern analysis and machine intelligence.
- [8] Jardine, A.K.S., Lin, D., Banjevic, D., (2006). "A review on machinery diagnostics and prognostics implementing condition-based maintenance". Mechanical Systems and Signal Processing; 20, p. 1483-1510.
- [9] M. Ageno, «Scribd.» [Online]. Available: www.scribd.com. [Accessed on 15th April 2013].
- [10] Maere, T., Villez, K., Marsili Libelli, S., Naessens, W., Nopens, I. (2012). "Monitoring the condition of membrane bioreactors with a combined PCA - FC algorithm". IWA World Water Congress, Busan, Corea del Sud, 16-21 settembre 2012, IWA Publishing, pp. 1-10.
- [11] MIMOSA, Version 3.3.1 Specification, Production Release, 2010, www.mimosa.org
- [12] NIST, 2011, <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>
- [13] Roberts, S.W. (1966). "A Comparison of Some Control Chart Procedures". Technometrics, pp. 411-430.
- [14] Roberts, S.W. (1959). "Control Chart Tests Based on Geometric Moving Averages". Technometrics, pp. 239-250.

Investigating the Effect of Water Contamination on Gearbox Lubrication based on Motor Control Data from a Sensorless Drive

Samieh Abusaad
Centre for Efficiency
and Performance
Engineering,
University of
Huddersfield,
Huddersfield, U.K.

Ahmed Benghozzi
Centre for Efficiency
and Performance
Engineering,
University of
Huddersfield,
Huddersfield, U.K.

Khaldoon F.
Brethee
Centre for Efficiency
and Performance
Engineering,
University of
Huddersfield,
Huddersfield, U.K.

Fengshou Gu
Centre for Efficiency
and Performance
Engineering,
University of
Huddersfield,
Huddersfield, U.K.
f.gu@hud.ac.uk

Andrew Ball
Centre for Efficiency
and Performance
Engineering,
University of
Huddersfield,
Huddersfield, U.K.
a.ball@hud.ac.uk

ABSTRACT

Water is one of the most significant destructive contaminations to lubricants, which in turn lead to more power consumption and early damage to rotating machines. This study explores the effect of water contents in gearbox lube oil on the responses of electrical supply parameters. A two stage gearbox based mechanical transmission system driven by a sensorless variable speed drive (VSD) is utilised to investigate experimentally any measurable changes in these signals that can be correlated with water contamination levels. Results show that the supply parameters obtained from both external measurements and the VSD control data can be correlated to the contamination levels of oil with water and hence can be based on for an instant diagnosis of water contamination. Particularly, the voltage and hence the power responses are more sensitive to the water contents than that of current because the VSD regulates more the voltage to adapt the small load changes due to the water induced lubrication degradation. Simultaneously, vibration also shows changes, which agree with that of power supply parameters.

Keywords

Sensorless VSD, Oil quality, Water in Oil Contamination, Motor Current Signature, Condition Monitoring,

1. INTRODUCTION

The main function of any lubricating oil is minimizing the friction between rotary and fixed machine parts, consequently reducing wear and temperature, which lead to high machine efficiency [14]. Deterioration of lube oil is often one of the main causes of machine breakdowns. A study conducted by Rabinowicz [11] shows that corrosion due to water in oil causes about 20% of all damages to mechanical equipment. Hence, water contamination detection has attracted high attention as to avoid catastrophic failures and increase the life of equipment. Main lubricant problems are from the contamination with water and dust. Corrosion due to water in oil, existence of incorrect additional fluids and failure of the oil itself due to temperature and heavy operational conditions are the most common issues from lubricant degradation [15].

The efficiency, lifetime and reliability of gearboxes and rotary systems are subjected to the quality of lubricating oil used. Therefore, it is important to monitor the condition of lubrication

oil to ascertain any degradation is progressing within the oil. The common technique used in industry for monitoring the oil quality is offline chemical and physical analysis. Samples of lube is normally taken every certain period and sent to chemical and physical laboratories for analysis and reporting the oil condition. Obviously, this approach usually is at high cost and cannot provide results timely

The survey in [9] has presented a review of lubricating oils condition monitoring techniques. They found that viscosity based equipment have better lubrication degradation feature coverage and lower cost. They also stated that oil viscosity is the most useful performance parameter that can reflect oil health and has been used as a standard feature to monitor the lubrication oil status.

Different new techniques have been developed to monitor the quality of oil online. For instance, work in [8] developed an online lubrication oil health monitoring and prognosis technique. Kinematic viscosity and dielectric constant physical simulation of the lubrication oil degradation due to water and the particle filtering algorithm is utilised. However, viscosity and dielectric sensors are required to implement the suggested technique.

In reference [1] an infrared (IR) sensing technique was suggested for online oil quality condition monitoring. The sensor is based on the evaluation of the oxidation number from the absorbed IR within a selected fixed narrowband IR filters. Additionally, the viscosity index (VI) and the base number (BN) of lube oils have been used to evaluating the oil quality in [10]. A multivariate calibration scheme has been used to determine the VI and the BN for the motor oils by applying the partial least squares (PLS-1) to the FTIR data.

Vibration signals together with advanced lubrication oil modelling schemes have been employed in [4] to provide corrosion related faults due to lubrication quality in aerospace gearboxes. However the performance of developed technique was not yet fully tested in terms of validating the reliability and accuracy.

The relationship between oil condition, i.e. lubricant viscosity and vibration signature was studied in [15], in which it is stated that some vibration features can be correlated with the oil viscosity. However the specified features have been failed in many experimental conditions.

Review the literature shows that the detection of gearboxes oil quality degradation, i.e. water contamination, is not well explored. Most techniques employed in industry are by offline monitoring, where the actual state of the lube cannot be determined instantaneously. Additionally, the available online techniques are not always applicable and expensive to apply. According to the research by [2], wind turbine and all other gearboxes application industries spend a considerable amount of money to solve problems and damages due to lubricating oil contamination.

In addition, the effects of water contamination levels has not been studied on both power supply parameters such as terminal current and voltage which have been shown to be effective measurements for monitoring mechanical faults[18][19]. Moreover, there is no known work so far studies the potential of using signatures from current, and control data for diagnosing lubrication problems.

Therefore, this study focuses on the development of water contamination monitoring methods based on electrical supply parameters. The possible effects of water contamination are reviewed firstly. Then a symmetric experimental investigation was carried out based on an industrial gearbox system running with different water contents. Through direct comparisons of different supply parameters between different water contents the diagnostic capability of these supply parameters is examined in line with corresponding vibration responses.

2. EFFECT OF WATER-IN-OIL ON GEARBOX TRANSMISSION SYSTEMS

Water in oil is the most destructive contaminant when mixed with lubricating oils. It causes significant chemical and physical changes [2]. Details on the negative consequences of water contamination in gearboxes oil can be found in [6] and [15]. The studies [3] and [20] shows that the most significant and immediate consequence is the clear changes in lubricant viscosity. It means that the dynamic and static behaviour of gearbox transmission systems will also change correspondingly.

2.1 Dynamic Effect

Based on gear lubrication mechanisms, the increase of viscosity will maintain more oil on the surfaces of meshing teeth and hence likely to form thicker hydrodynamic films, leading to a reduction of frictional force. On the other hand, the decrease of viscosity will lead higher friction. Furthermore, as the occurrence of tooth meshing is a time-varying process according to mainly the meshing period, it is expected that the change of frictional forces will alter the vibration responses of gearbox at meshing frequencies and their high order harmonics. As shown in [15], vibration spectrum is strongly related to the lubricant viscosity in a high frequency range from 2300 Hz to 7000Hz based on a 0.34kW gearbox.

In addition, it is also likely that this change of tribological behaviour will affect the dynamic responses of components associated with gear shaft rotation frequencies. These components exist because of evitable gear manufacturing errors such as misalignments and accumulative pitch deviations, which cause non-uniform meshing process and eventually lead to the change of the dynamic responses at shaft frequencies.

2.2 Static Effect

Obviously, the change of tribological behaviour will also lead to changes of average friction losses. It means that the overall load of the system will be different for different water contents.

Besides, according to the study in [17], the density, viscosity and oil squeezing will also show considerable losses of power, which will reflect on static measurements such as average supply current, voltage and thus the power.

2.3 Detection based on Sensorless Drive

The dynamic effect can be reflected by vibration measurements. However, the vibration may not so sensitive to the static effect. In contrast, previous studies [18] [19] have shown that both dynamic effects and static effects occurring on downstream machines can be observed in motor current signals. Specifically, the dynamic effect can be represented by the sideband components at [16]:

$$f_x = f_s \pm kf_r \quad k=1,2,3,\dots \quad (1)$$

where f_x denotes fault frequency, f_s is supply frequency and f_r is rotor frequency, whereas the amplitude change at the supply frequency can be based to examine the static effects.

Nowadays, VSDs are used more and more in industry for achieving more efficient production. When VSD is under sensorless mode, the motor speed at the setpoint will be maintained relatively stable by compensating any changes due to load disturbances through more advanced control algorithms, which may give more chances to observe these two effects, rather than just using the feature shown in formula (1) based on the high slippage changes under direct voltage/Hz control methods.

As shown in Figure 1, when a speed deviation due to disturbances from the reference is detected by the driver, the speed comparator sends a speed error to the speed control loop which motivates the controller of the torque current component (i_d). The output signal from the i_d current controller sets the reference torque required for compensating the effect of the disturbances. The torque reference is then compared with the estimated torque by the torque controller, which outputs the required torque voltage component (v_q) to the pulse width modulator (PWM). In the meantime, the field current component (i_d) control loop sets the required flux (ψ). The flux normally kept fixed at the nominal motor value when speed lower than the rated motor speed. Thought, when the induction motor operates at or higher the rated speed a field weakening is required to prevent exceeding the voltage limit of both the drive and the motor. The field weakening is performed at the flux control loop by a fraction of the rotor angular frequency ($1/\omega_r$). This control process therefore shows that it is possible to observe the changes of oil properties through exploring different control variables.

It is worth mentioning that the sensorless VSDs estimates both motor speed and flux by measuring terminal supply parameters without the need for additional speed measuring device.

3. TEST FACILITY AND PROCEDURES

The test facility employed for this study consists of a mechanical system and a control system. As shown in Figure 2 the mechanical system includes of a 15kW AC induction motor (IM) as the prime driver, two back-to-back two stage helical gearboxes for coupling the AC motor with a DC load generator using flexible spider

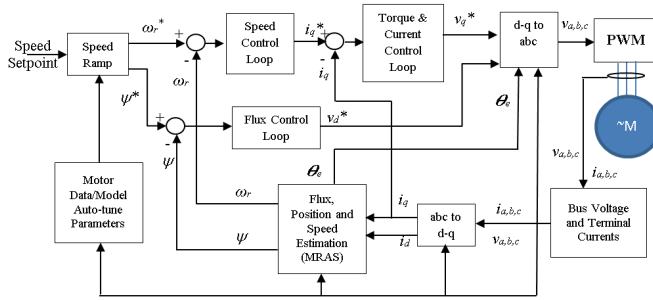


Figure 1. General structure of sensorless VSD

rubber couplings. The first gearbox operates as a speed reducer while the second is a speed increaser so that the system maintains sufficient speed for the DC load generator to produce sufficient load to the AC motor through the two gearboxes. The control system consists of a programmable logic controller (PLC) for setting up different speed-load profiles specified by operators, an AC VSD that can be set either to a sensorless flux vector control mode or V/Hz mode for adjusting the speed of the system, a DC variable speed drive providing a controlled load to the AC motor by regulating the torque of the DC load generator.

All data sets from both data acquisition systems are stored in a PC for post processing and analysis in the Matlab environment.

The dynamic data is used for both evaluating the performance of the conventional analysis methods for detecting water contamination problems, and also for benchmarking the developed scheme from the static data.

Experimental studies were carried out by adding different amount of tap water to the lube oil in the first gearbox (GB1), which is a common industrial gearbox with a speed ratio of 3.6 and a power

transmission of 10kW at 1450rpm. After a baseline (BL) test when the gearbox was filled with new EP320 gear oil, four incremental water contents: 4.0kppm, 7.0kppm, 30.0kppm and 60.0kppm were tested one by one, which correspond to 0.4%, 0.7%, 3% and 6% water in oil. The first two water contents are below the recommended level [7] and the last two are above the level, which allows the variation of different measurements underlying to be examined in a wide range for defining their corresponding detection performances.

During the test, the rig operated under five increment load settings: 0%, 25%, 50%, 75% and 100% of the gearbox when it is at its full speed of 1460 rpm, attempting to examine the detection performance under variable load operations which is common scenarios of real applications. Each load setting was for a period of two minutes and changed to next one automatically by the PLC controller. In total, this load cycle lasts 10 minutes. In addition, the VSD was set under sensorless control mode for evaluating the detection capability under this particular mode.

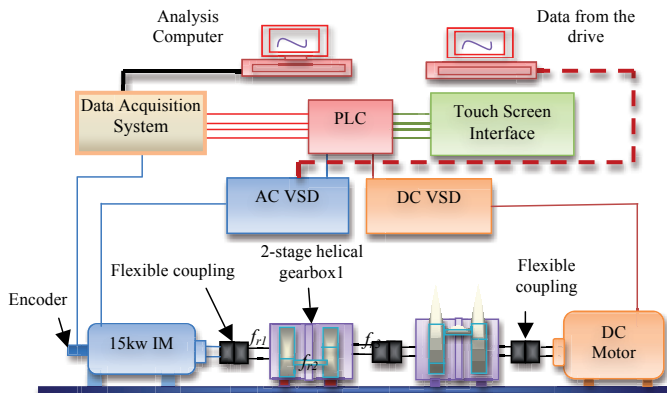


Figure 2. Schematic of test system

To ensure the data quality for reliable comparison the load cycle ran consecutively seven times for each of the water contents. During these repeating operations, the lube temperature in GB1 was observed on-line and reached to 51°C-52°C from the room temperature when the system operating parameters also became stabilised.

By using an automated acquisition procedure based on time advancement, 30 seconds of dynamic data were collected for every load setting. In the meantime, the static data from the VSD were also logged for the entire load cycle. Moreover, oil samples were taken after each incremental water test for measuring the viscosity of the water contained oil.

4. RESULTS and DISCUSSION

The data sets collected were processed by necessary schemes including spectrum analysis, ensemble average and time synchronous average (TSA) to obtain reliable feature parameters for implementing an effective detection of the water contents added to the lube oil in GB1.

4.1 Measurement Validation

To examine the data quality and test reliability, signals including data from the lubricant temperature, speed, current and voltage are processed to obtain their static feature values. Figure 3 shows

these key measurements against testing run numbers for 100% load. It can be seen in Figure 3 (a) that the temperature of lubricant in GB1 increases gradually and reaches stable status at 6th test when the test system becomes stabilised. Moreover, temperature values between different water contents show little difference, showing that these tests were carried out with good repeatability. This repeatability can be also seen from the phase current and the motor speed measurements, which are presented in Figure 3 (d) and (e) respectively. In addition, both of these two parameters show gradual decrease behaviours with the test number because the transient effects including both the higher frictional influences of the test rig and larger resistance values of the electric and electronic components employed in the motors and the VSD control system when they are under lower temperature operation. This transient effect can be seen to be much stronger when the load and speed characteristics in Figure 3 (f) are explored. When the system operates under low temperatures during the 1st test run, the VSD has poorer performance in maintaining the speed of the system at the setting points, which leads to higher speed under the higher load. On the other hand, when the system reaches its stable value the VSD is able to control the speed with higher accuracy under different load settings, as shown by the speed results for the 7th run.

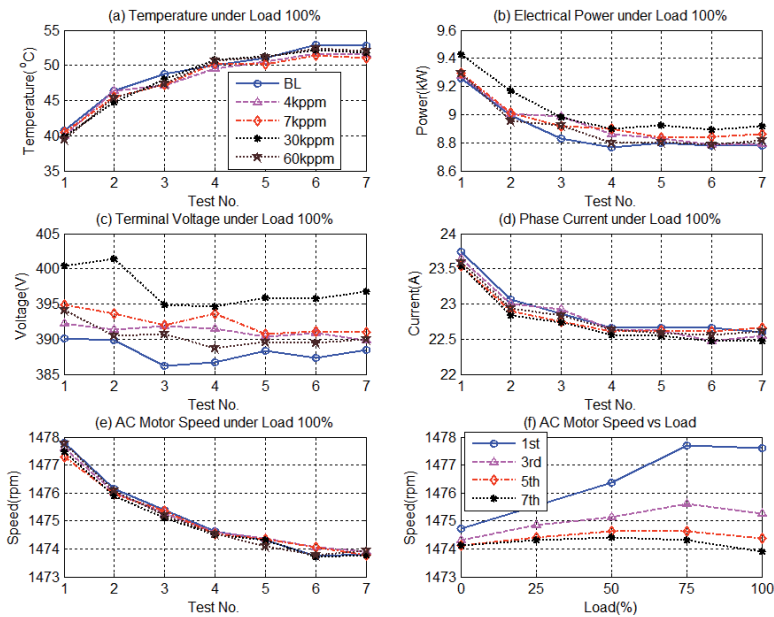


Figure 2. Key measurements variation with test run numbers and water contents

However, the terminal voltage in Figure 3 (c) shows relatively stable over different testing run numbers. Moreover, it shows

clear difference between different water contents, showing that it may sensitive to the changes due to lubricant properties. The

similar changes between different water contents can also be observed in the electrical power shown in Figure 3 (b) as the power is calculated by multiplying the voltage with the current directly to show the overall power consumption of the system.

Based on these test observations, it can be concluded that the measurements from last test run numbers have less transient effect and be more stable for examining the effect of water contents more accurately. Moreover, it has exposed that the VSD seems to adjust supply voltages more to adapt the changes due to lubrication conditions.

4.2 Effect on Viscosity

As the most important performance parameter of lubricating oils, the viscosity was measured using a Kinexus pro+ rheometer for each water content test and taken as references for representing lubrication deterioration and for comparing the detection performance using supply parameters. Figure 4 shows the representative viscosity values at 50°C and 55°C, which are close to the temperature range where the gearbox is under stable operating. The viscosity shows a slight decrease from that of the base oil for small portion of water content. Then it shows a monotonic increase with water contents, which is slightly different to the result published in [3] and [20]. The effect of very small amount of water decreases the viscosity, up to a certain level i.e. at around 4kppm where it starts to increase due to the interaction of the water droplets. The viscosity starts to be around the amount of that of the base oil when water content is about 60kppm. Nevertheless, this measurement confirms that the water content was added effectively according to the test design.

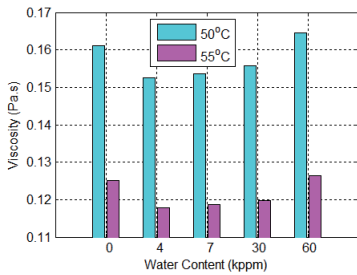


Figure 4. Variation of viscosity with different water contents

In addition, Figure 4 also shows that temperature has more significant influence on the viscosity i.e. the higher temperature corresponding to lower viscosity. This also explains that the electrical power decreases gradually with test run numbers presented in Figure 3.

4.3 Effect on Power Supply Parameters

Further analysis of the dynamic current and voltage signals has found that the dynamic responses such as sidebands of the supply frequency has shown little constant changes to the water contents because of high moment of inertia of the test system. Therefore, only the static parameters are discussed further.

Figure 5 presents a comparison of power supply parameters for the lubricants with different water contents. These results were

obtained by averaging the results from the 6th and 7th testing runs under 100% load in order to give a more reliable result. The terminal voltages in Figure 5(a) and their corresponding power values have shown good correlation with the changes in viscosity. For the base value, the voltage and hence power show the lowest value, reflecting the good performance of the base oil in reducing the friction because of its relatively high viscosity. For the cases of small water added to the oil such as 4kppm, 7kppm and 30ppm, the viscosity values reduce and lead to poorer lubrication. It means that they need more power to overcome the friction loss to maintain the constant speed operation. Therefore, the electrical power shows higher values for these cases. However, the 30kppm case shows the highest power consumption due to other effects such as more oil squeezing. For the 60kppm case, the viscosity is higher than that of the base oil and the effect of viscosity may higher than that of oil squeezing, which lead to that the power consumption goes lower.

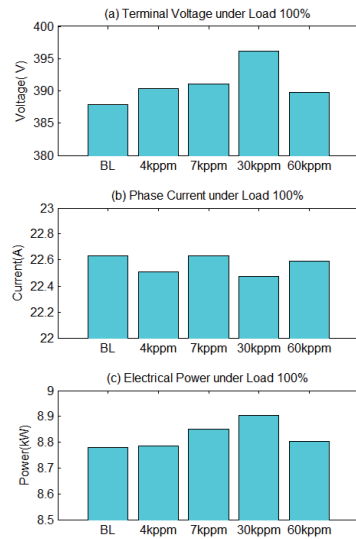


Figure 5. Temperature comparison during 7th test

However, because of the regulation of VSD, the power consumption characteristics are not reflected by the current measurements. Instead, the current measurement shows similar value for the different water contents as shown in Figure 5 (b). Therefore, it may not be suitable for detecting the water contamination.

In addition, the control data obtained directly from the VSD also exhibits the same characteristics shown in Figure 5.

These results show that the electrical measurements from a VSD can be used to separate different levels of water contamination. However, both the voltage and current need to be measured in order to obtain the power for a reliable separation even though the voltage measurement shows better performance of detection.

4.4 Effect on Vibration Responses

To show the dynamic effect of the water content, the vibration signals are processed to obtain root mean squared values (RMS) to represent overall level of vibration responses for different water contents. Figure 6 shows the averaged results of the last two test runs. It can be seen that the vibration levels for different water content are higher than that of the base line, which is consistent with the results from the static electrical parameters in that the water degrade the lubrication performances and cause more friction between the surfaces of meshing teeth. Especially, compared with the vibration from GB2, which shows relatively unchanged between different tested cases, the vibration changes of GB1 are clearly significant and can be used to indicate the water contamination.

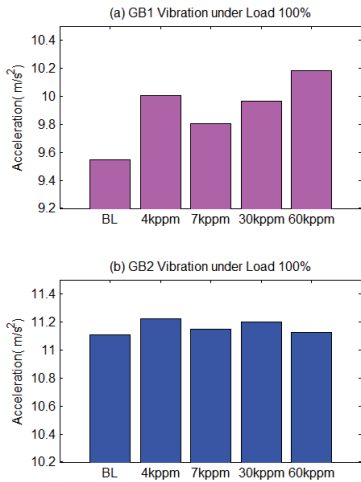


Figure 6. Vibration spectra under different loads and cases

To understand the vibration changes further, the vibration signals are applied with a time synchronous average (TSA) procedure and subsequent order spectrum analyse to suppress noise influences which are not time aligned to the second shaft and hence to obtain vibration which are more associated with gear meshing dynamics. Figure 7 (a) and (b) shows the amplitudes of vibration components at the two mesh frequencies of the two stage gearbox, which is the average amplitude up to the 10th high order harmonics. The vibration at the first mesh frequency which is the high speed stage with high load shows higher amplitudes compared with the baseline and exhibit similar behaviour with that of the static power parameters. The vibration at the second mesh frequency which is the low speed stage with high load also shows higher amplitudes compared with the baseline and that of the higher speed stage and exhibit more similar behaviour with that of the static power parameters. This shows that the high load stage is more influenced by the change in lubrication, which is consistent that the hydrodynamic oil film is less helped because the relative velocity between meshing tooth surfaces changes its direction at the pitch line. In addition, these vibration changes are

consistent with that in [15]. Furthermore, the change of vibration responses provide further support that the power supply parameters can reflect the changes in gear lubrication due to water contamination.

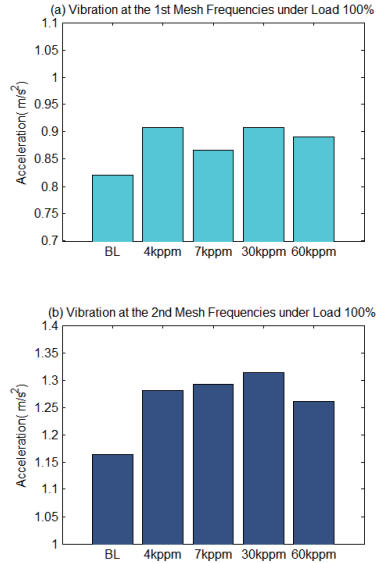


Figure 7. Vibration at mesh frequencies

5. CONCLUSION

In this paper a new cost effective technique is presented to detect gearbox lubricant contamination with water based on power supply parameters. Experimental results shows the viscosity reduces when small portions of water (<30kppm) added to the gear oil while the large portion of water increases the varicosity. Based on a 10kW gearbox rig, this contamination can be measurable using the static power supply parameters including the voltage and power. However, the current values show little change because the VSD adjusts the electrical power supply to adapt to the changes due to the frictional load through regulating more the voltage than the current.

In addition, the results have been supported by vibration response analysis in that the vibration responses at high load stage show much similar change to that of power supply parameters due to the water contents.

Moreover, this study shows that both power supply parameters and vibration responses can be effective measurements to be based on for performing the instant diagnosis of water contamination so as to prevent further damages to gearboxes running with water contents.

6. REFERENCES

- [1] Agoston OVE, C. Schneidhofer, N. Dorr, and B. Jakoby (2008). "A concept of an infrared sensor system for oil condition monitoring". *Elektrotechnik & Informationstechnik* 125(3), pp. 71–75.
- [2] Bernard J.C. (2004). "Predictive & Preventive Maintenance in Lubricating Systems". *DES-CASE Europe sarl. Available at: http://www.descase-europe.com/uploads/1256143218_lubricated_systems.pdf*
- [3] Jakoby, B. and Vellekoop, M.J. (2004). "Physical sensors for water-in-oil emulsions". *Sensors and Actuators Published by Elsevier B.V.*, 110, pp. 28–32
- [4] Byington, C. S., Brewer, R. C., Meyer, T., and Amin, S. R. (2007). "Gearbox Corrosion Prediction via Oil Condition Sensing and Model Fusion," *STLE 62nd Annual Meeting and Exhibition, Philadelphia*.
- [5] Campbell W S; Jacopini J; 1994, "Suction Roll Pilot Bearing Failure Due to Inadequate Lubrication", *Tappi Journal* , 77(12), pp. 233-236
- [6] Echin, A. I., G. T. Novosartov, T. B. Kondrat'eva. (1983). "Effect of water on lubricating properties of synthetic oils". *Chemistry and Technology of Fuels and Oils*, 19(2), pp. 80-83.
- [7] Jakoby, B., and Vellekoop, M.J. (2004). "Physical Sensors for Water-In-Oil Emulsions". *Sensors and Actuators A*, vol. 110, pp. 28 - 32.
- [8] Junda Zhu, David He, and Eric Bechhofer. (2013). "Survey of Lubrication Oil Condition Monitoring, Diagnostics, and Prognostics Techniques and Systems" *Journal of Chemical Science and Technology*. 2 (3), pp. 100-115.
- [9] Junda Zhu, Jae M. Yoon, David He, Yongzhi Qu, and Eric Bechhofer. (2013). "Lubrication Oil Condition Monitoring and Remaining Useful Life Prediction with Particle Filtering". *International Journal of Prognostics and Health Management* ISSN 2153-2648.
- [10] M.A.Al-Ghouthi, Y. S.Al-Degs and M. Amer. (2010). "Application of chemometrics and FTIR for determination of viscosity index and base number of motor oils". *Talanta*, 81, pp. 1096 – 1101.
- [11] Mike Day Pall and Mika Vesala. (2009). "Setting Control Limits for Water Contamination in Hydraulic and Lube Systems". *The 10th Scandinavian International Conference on Fluid Power*. Available at: <http://ebookbrowse.net/the-10th-scandinavian-international-conference-on-fluid-power-pdf-d55044223>.
- [12] Paladugu, A., and Chowdhury, B. (2007) 'Sensorless control of inverter-fed induction motor drives' *Electric Power Systems Research*, 77 pp.619–629
- [13] R.E. Cantley. "The Effect of Water in Lubricating Oil on Bearing Fatigue Life". *ASLE Transactions, American Society of Lubrication Engineers*, 20(3), pp. 244-248; from a presentation at the 31st Annual ASLE Meeting, Philadelphia, Penn.
- [14] Rajeev R. and Harsh Kumar D. (2013). "Effect of Lube Oil Temperature on Turbine Shaft Vibration". *International Journal of Mechanical Engineering and Robotics Research (IJMERR)*, 2(2), pp. 2278 – 0149
- [15] Rui Gomes de Almeida and Linilson Rodrigues Padovese. (2003). "Characterization of oil viscosity alterations in a gearbox through vibration signal analysis". *17th International Congress of Mechanical Engineering (COBEM2003)*. Available at: <http://www.abcm.org.br/pt/wp-content/anais/cobem/2003/html/pdf/COB03-0555.pdf>
- [16] Schoen, R. R. and Habetler, T. G. 1995. Effects of time-varying loads on rotor fault detection in induction machines. *IEEE Trans. Ind. Appl.*, Vol. 31, pp. 900–906.
- [17] Gorla, C., Concli, F., Stahl, K., Hohnet, B. Klaus, M. Schultheiß, H. and Stemplinger, J. (2012) "CFD Simulations of Splash Losses of a Gearbox, , *Advances in Tribology*". *Hindawi Publishing Corporation*, Volume 2012.
- [18] Tian, X., Feng, G. Albraik, A., Gu, F., and Ball, A.D. (2014), "The investigation of motor current signals from a centrifugal pump for fault diagnosis". *Comadem 2014*, Brisbane, Australia.
- [19] Abusaad, S., Benghuzzi, A., Shao, Y., Gu, Fengshou and Ball, A., 2013. "Utilizing Data from a Sensorless AC Variable Speed Drive for Detecting Mechanical Misalignments". *Key Engineering Materials*, 569. pp. 465-472. ISSN 1013-9826.
- [20] Johnsen, E. and Ronningsenb P. (2003) "Viscosity of 'live' water-in-crude-oil emulsions: experimental work and validation of correlations". *Journal of Petroleum Science and Engineering* 38 , pp. 23– 36

eMaintenance 2B

Identification of Factors affecting Human performance in Mining Maintenance tasks

Mojgan Aalipour

Division of operation and Maintenance,
Luleå University of Technology,
+46920492859

Mojgan.alipore@ltu.se

Sarbjeet Singh,

Division of operation and Maintenance,
Luleå University of Technology,
+46920492812

Sarbjeet.singh@ltu.se

Abstract

This paper investigates the factors affecting human performance in maintenance task in mining sector. The objective is identify various factors and to classify them as driving (strong driving power and weak dependence) and dependent factors (weak driving power and strong dependence). The factors were identified through literature survey and are ranked using mean score of data questionnaire. The reliability of measures is pretested by applying Cronbach's alpha coefficient to responses to a questionnaire given to maintenance personnel. The interrelationships between human factors have been recognized by interpretive structural modeling (ISM). Further, these factors have been classified using matrice d'impacts croises-multiplication appliqué à un classement (MICMAC) analysing. This case study will figure out the factors affecting human performance for deriving maintenance management insights to improve productivity in the mining sector. Further, this understanding may be helpful in framing the policies and strategies for mining industry. Temperature, lighting, documentation, communication and fitness are driving factors. Moreover, Work layout, tools availability, complex tasks, time pressure, safety, boss decisions, training, fatigue and motivation have strong driving power as well as high dependencies and it comes under the category of linkage factors.

Keywords

Human Factors, Maintenance, Interpretive structural modeling, Micmac analysis

1. INTRODUCTION

The field of human factors and ergonomics is multidisciplinary which includes psychology, engineering, computer science, physiology, biomechanics, industrial design, anthropometry etc. [1, 2]. Various organizations have given different definitions of human factors and ergonomics. As per the International Ergonomics Association "Ergonomics (or human factors) is the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data and methods to design in order to optimize human well-being and overall system performance. Human factors contribute to the design and evaluation of tasks, jobs, products, environments and systems in order to make them compatible with the needs, abilities and limitations of people" [1]. Since the 1960's, significant reports have been documented in human factors related mining issues [3] and reflects that maintenance field is largely reliant on human activity and can effect performance and productivity [4]. Like other industrial sectors, the mining industry has also experienced the rapid application of modern technology and automated activities due to competitive global economy [5]. This advanced technology has obviously increased the complexity of tasks and time in maintenance field, which

significantly increased the human error and is considered as a root cause of many accidents and incidents [6]. Automation and modern technology in today's mines needs more efficient and competent employees as well [5]. Moreover, unfavourable working conditions and poor maintenance performance can have many consequences in different dimensions such as: effect on the health and safety of personnel, inconvenience and loss of production, loss of life, commercial effects, premature equipment failures, damage to plant and equipment, customer dissatisfaction or loss of business [4]. The objective of this paper is to identify and evaluate the factors affecting human performance and classify them as driving and dependent factors in mining maintenance context. The study can help maintenance management to identify the critical factors affecting human performance. Interpretive structural modelling (ISM) is a well-established methodology for identifying relationships among specific items. This methodology has been increasingly used by various researchers to represent the interrelationships among various factors. Interpretive structural modelling (ISM) was first proposed to analyse complex socioeconomic systems [24]. The basic idea is to use expert knowledge and experience to decompose a complex system into several subsystems and construct a multilevel structural model [7-8]. Many researchers [7-15] have used the ISM method in recent years to represent the interrelationships between various elements linked to the problem or issue. ISM has been used for policy analysis [25] for management research [12, 14] follow up ISM to develop a hierarchy of actions required to reach the future objectives of waste management in India. The structural analysis used in the study is basically a tool that structures the sharing of ideas. This type of analysis defines a system using a matrix which combines the components of the system. It has also been observed from the literature that MICMAC analysis [11-13] has been extensively used to identify and to analyze the variables according to their driving power and dependence power [13-15].

2. FACTORS AFFECTING HUMAN PERFORMANCE IN MINING SECTOR

In this case study 31 factors have been identified from literature which may have effect on human performance in mining maintenance tasks. The factors were classified under three groups, namely organizational factors, individual factors and job and workplace factors.

2.1 Organizational Factors

Organizational factors are representative part of each organization that illustrates the structure and systems, such as leadership, structure, process and procedure, communication etc. In this paper, we address the organisational factors from maintenance perspective such as, duties, responsibility and authority, documentation (rules & procedures, manuals), communication, boss decisions (untimely, wrong), safety system (instruments, cautions & warnings), contract, salary

and breaks [6]. In order to improve human performance “management must define the structure, hierarchy and lines of reporting. It must ensure that duties, responsibilities and authority of all personnel are well defined and communicated.” [14]. Moreover, set of rules, procedures, instructions and standards must be clear for improving quality products and safe job [14, 15]. Communication between managers and personnel must be defined [14] as poor communication may cause a failure in the system due to miscommunication. Supervisors play a major role by effectively carrying administrative and technical tasks to improve performance, on time task completing, effective communication between employees and managers, work layout [16]. Furthermore, it has also been observed from literature that human performance gets affected by safety systems within the work place. These systems consist of assessment of units, inspection, and control of hazards and risks which affect workplace health and safety. Unsafe working atmosphere can increase the fear of accidents and affect performance [17]. Moreover, in order to motivate employees companies should offer contracts giving an indication of job security besides addressing duties, responsibilities, employment conditions, salary and rights [18].

2.2. Individual Factors

Individual factors such as: education, training, fitness for duty, skill, motivation, stress and fatigue play significant role in human performance. Effective training program makes ability to do the task in better way and minimize the time to take decision. Moreover, fitness-for-duty assists workers in maintaining physical, mental and emotional capacity to perform given tasks [19]. Skill is a personal quality and the reliability of performing the job stress plays important role in the workplace. The following figure shows maximum human performance effectiveness happens at moderate levels of stress and not at low or high stress. By increasing stress levels, performance will decrease and it increase probability of errors in maintenance activities and causes other types of psychological stress such as: anxiety, fear [20].

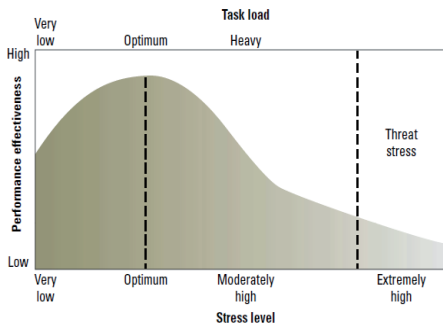


Figure 1. Human performance effectiveness versus stress curve [21]

2.3. Job & Workplace Factors

Workplace design incorporates designing the task based on the ability and limitations of people in order to improve human performance, prevent overload, pressure and to match with task requirements and health and safety [22]. It includes tools design, time pressure, task complexity, repetitive tasks, shift work and environment factors: noise, humidity, vibration, lighting, dust, temperature, smelly fumes, and

slippery floors. Poor tools design, time pressure can affect safety, workload, performance, and can cause errors. Task complexity is one of the significant factors affecting human reliability [15]. Poor environmental workplace with high noise levels, high or low temperatures, high humidity, and poor lighting and ventilation, smelly fumes and slippery floors can effect performance, motivation and health of personnel and increase the likelihood of human error [4].

3. RESEARCH METHODOLOGY

The case study was conducted in a mine based in a northern part of Sweden, however, due to confidential reasons; information related to company has been masked. The data collected through a questionnaire survey involving the participation of 18 employees dealing with maintenance of mining machinery. The respondents were male with the average 16 years’ experience. An interview was also conducted with 4 of experts who were maintenance supervisors. In addition, 2 academics with mining experience have been interviewed. A questionnaire survey was conducted to rank the factors affecting human performance in mining maintenance activities. In the present case study, MICMAC analysis is implemented for identifying the factors which are influential, dependent and essential to the evolution of the system. MICMAC analysis is the phase in which the key factors were identified those essential to the system’s development. The main advantage of using this analysis is that it motivates thought process and generates ideas among expert members. Even if being a subjective approach it encourages the outer-intuitive aspects of how a system works. In all, 31 factors were identified from literature review. In the questionnaire, maintenance personnel were asked to designate the significance of 31 factors on a 5-point Likert scale. On this scale, “1” and “5” corresponded to “not at all” and “very often/much” respectively. The questions includes; how much documentation and poor communication influences the task, how much untimely, incorrect or immediate decisions affects human performance etc. The questionnaires were analysed using SAS - JMP11 software. Cronbach’s alpha coefficients were applied to the responses to determine reliability. In addition, the factors are ranked based on mean scores value and standard deviation which presented in the Table 1. Standard deviation measures variability within a distribution. The reliability of data will specified with lower standard deviation [23]. For validation of the questions standard deviation of data have been calculated.

3.1. Interpretive structural modeling (ISM):

A “leads to” contextual relationship was chosen to analyse the relationship among the factors. In the process of developing SSIM (Structural Self Interactive Matrix), the symbols (V, A, X, O) were used to denote the direction of the relationship between two factors (i and j) (Table 3). V is the relation from factor i to factor j (i.e. if factor i influences or reaches to factor j), A is the relation from factor j to factor i (i.e. if factor j reaches to factor i), X is used for both direction relations (i.e. if factors i and j reach to each other), and O indicates no relation between two factors (i.e. if factors i and j are unrelated) (Table 2).

Table 1. statistical analysis

Factor Area		Factors	Mean Score	Standard Deviation	Mean of Standard Deviation	Mean of Mean Score		
Job & Workplace Factors	Physical Environment	Temperature	4.00	0.89	1.12	3.67		
		Lighting	3.88	1.15				
		Vibration	3.81	0.98				
		Smelly Fumes	3.81	1.05				
		Slippery Floors	3.75	1.24				
		Dust	3.50	1.03				
		Noise	3.38	1.26				
	Workplace & The Equipment used	Humidity	3.25	1.34	1.12	3.50		
		Work layout	3.94	1.18				
		Tools Availability	3.75	1.24				
	Job & Demands on the worker	Tools Design	3.63	1.26	1.23	3.77		
		Complex Tasks	3.38	0.96				
		Time Pressure	3.26	0.85				
		Repetitive Tasks	3.25	1.13				
Work Overload		3.00	0.89					
Shift Work		2.38	1.15					
Organizational Factors	Documentation	3.69	1.25	1.16	3.18			
	Communication	3.38	1.31					
	Safety System	3.31	1.08					
	Boss Decisions	3.25	1.24					
	Duties& Responsibilities	3.25	1.18					
	Salary	2.94	0.93					
	Contract	2.94	1.29					
	Breaks	2.75	1.00					
	Individual Factors	Training	3.50			1.15	1.08	3.0
		Fatigue	3.31			1.01		
Motivation		3.25	1.24					
Fitness		3.19	1.05					
skill		3.00	1.03					
Education		2.88	1.09					
Stress		2.88	0.96					

Table 2 . Direction of factor's relationship

Condition	Factor's relationship	symbols
1	$i \rightarrow j$	V
2	$j \rightarrow i$	A
3	$i \leftrightarrow j$	X
4	$I X j$	O

Based on these contextual relationships the SSIM is developed. A reachability matrix is a binary matrix (1, 0). The structural self-interactive matrix is transformed into initial reachability matrix by substituting V, A, X and O by 1 and 0. After developing initial reachability matrix, the concept of transitivity is used to fill some of the cells of the initial reachability matrix by inference. After incorporating the transitivity concept, the final reachability matrix is obtained. The final reachability matrix indicates the driving power and dependence of each factor (Table 4). Dependence is the total number of variables (including itself) which may be impacting a factor. The driving power for each variable is the total number of variables (including itself), which it may impact [7].

4. RESULTS

In this case, the value of Cronbach's alpha coefficients for each questions is ≥ 0.95 . Statistical data analysis consists of mean score, standard deviation which are reflected in Table 1. High standard deviation indicates large variations in the data, and low standard deviation symptom for small variations in the data. In our case study the range of standard deviation are

located between 0.85 and 1.34, hence the questions are reliable. Based on questionnaires result, 14 factors with high mean score have chosen from 31 factors. These are highlighted in the Table 1.It is pertinent to mention that fourteen factors identified from literature are mutually influential besides affecting human performance. The factors have been ranked based on mean score. Furthermore, the factors have been classified using MICMAC analysis technique. MICMAC analysis refers to Matrice d'Impacts Croisés Multiplication Appliquée à un Classement. It involves the development of a cluster to classify factors as driving or dependent. The factors affecting the performance of human operators in maintenance tasks are classified as: autonomous factors (weak driving power and weak dependence), linkage factors (strong driving power as well as strong dependence), dependent factors (weak driving power but strong dependence) and independent factors (strong driving power but weak dependence power). The drive power-dependence power diagram is shown in Figure 2. Temperature, lighting, documentation, communication and fitness are driving factors. In other words; they have strong driving power and weak dependency on other factors. They may be treated as the key factors affecting human performance in mining maintenance tasks. Work layout, tools availability, complex tasks, time pressure, safety, boss decisions, training, fatigue and motivation have strong driving power as well as high dependencies and it comes under the category of linkage factors. If these factors are accommodated, there will be a positive influence on maintenance with a reduction in human error.

Table 3. Structural self-interactive matrix (SSIM)

NO	SSIM	14 Fitness	13 Motivation	12 Fatigue	11 Training	10 Boss Decisions	9 Safety	8 Communication	7 Documentation	6 Time pressure	5 Complex Tasks	4 Tools availability	3 Lighting	2 Work layout	1 Temperature
1	Temperature	O	V	V	O	O	O	O	O	O	O	O	O	O	
2	Work layout	O	V	V	O	A	X	O	O	O	O	O	V	V	
3	Lighting	O	V	V	O	O	O	O	O	O	O	O			
4	Tools availability	O	O	V	O	A	O	O	O	O	O				
5	Complex Tasks	O	V	V	A	X	O	A	A	O					
6	Time pressure	O	V	V	O	X	V	O							
7	Documentation	O	O	O	O	X	A	V							
8	Communication	O	V	O	O	X	O								
9	Safety	V	V	A	A	X									
10	Boss Decisions	V	X	V	X										
11	Training	O	X	A											
12	Fatigue	A	A												
13	Motivation	A													

Table 4. Initial and final Reachability Matrix

No	Reachability Matrix	Temperature	Worklayout	Lighting	Tools availability	Complex Tasks	Time pressure	Documentation	Communication	Safety	Boss Decisions	Training	Fatigue	Motivation	Fitness	Driver Rank
1	Temperature	1	0	0	0	1	1	0	0	1*	1*	1*	1	1	0	8
2	Worklayout	0	1	1	1	1	1	0	0	1	1*	1*	1	1	0	10
3	Lighting	0	1	1	1*	1*	1*	0	0	1*	0	1*	1	1	0	9
4	Tools availability	0	1	1*	1	1	1	0	0	1	1*	1*	1	1	0	10
5	Complex Tasks	0	1*	0	1*	1	1	1*	1*	1	1	1*	1	1	1*	12
6	Time pressure	0	1*	0	1*	1	1	1*	1*	1	1	1*	1	1	1*	12
7	Documentation	0	1*	0	0	1	1*	1	1	1	1*	1	1*	1	0	10
8	Communication	0	1*	0	1*	1	1*	1	1	1*	1	1*	1*	1	1*	12
9	Safety	0	1	1*	1*	1*	1*	1*	1	1	1*	1*	1*	1	1*	13
10	Boss Decisions	0	1	1*	1	1	1	1	1	1	1	1	1	1	1	13
11	Training	0	1*	0	0	1	1*	0	0	1	1*	1	1*	1	0	8
12	Fatigue	0	1*	0	0	1	1	0	0	1	1*	1	1	1*	0	8
13	Motivation	0	0	0	0	1	1	0	0	1*	1*	1	1	1	0	7
14	Fitness	0	1*	0	0	1*	1*	0	0	1	1*	1*	1	1	1	9
Dependence Rank		1	12	5	8	14	14	5	5	14	13	14	14	14	6	

Driving Power	14														
	13		Driving Factors						Linkage Factors					F10	F9
	12				F8										F5,F6
	11														
	10				F7			F4				F2			
	9				F3	F14									
	8	F1													F11,F12
	7														F13
	6														
	5		Autonomous Factors						Dependent						
	4														
	3														
	2														
	1														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Dependence Power															

Figure 2. MICMAC Analysis

5. DISCUSSIONS

Organizations should concentrate on preventing accidents, removing or decrease human failures and improving human performance and safety issues [22]. The rapid growth of technology in mining sector not only upgraded productivity but also brought a sequence of complex system, which are to be maintained by humans and not machines. It is an open secret that human factor has a significant role in all steps, particularly in maintenance. It therefore becomes essential for mining sector to know how to use human requirements to improve performance and prevent accidents through in all the process and systems [15]. It has been well known that well organizations should concentrate on prevent accidents, remove or decrease human failures and improve human performance and safety issues [22]. It has been well known that well trained, motivated, under no time pressure, with proper work layout and maintenance friendly tools designed, a maintenance crew can deliver good performance [24], but there is dire need to implement what is written in reports and documents. This case study successfully identified the factors effecting human performance. Further, it has also been observed from the analysis of questionnaire that the identified factors have very relevant effect on the performance of workers while carrying out maintenance task. Moreover, the factors affecting human performance in maintenance task in mining sector have also been ranked to make way for future studies.

6. CONCLUSIONS

Maintenance in the mining sector must be kept in decent working condition; many factors directly or indirectly result in a decline in human performance, causing errors in maintenance. This paper's objective is to identify and analyse the factors affecting the human performance in mining maintenance tasks. The driver power-dependence matrix sheds light on the relative importance of each factor and the interdependence among the factors. It can be concluded that the factors, such as, work layout, tools, complex tasks, training, boss decisions, time pressure, fatigue, safety, and motivation have been identified as driving factors (critical factors). According to MICMAC analysis these driving factors

will greatly influence human performance in mining maintenance tasks. It is therefore strongly recommended that the mining company should put more emphasis on these driving factors. Moreover, the linkage factors should also be further investigated so that their effect on human performance during maintenance task can be neutralized. The present case study can help maintenance management understand the interaction of factors affecting human performance in maintenance and assist in devising policies and guidelines for improving mining maintenance related tasks.

7. ACKNOWLEDGMENT

The authors would like to thank for the support of CAMM (Centre of Advanced Mining & Metallurgy) project in this research work.

8. REFERENCES

- [1] International Ergonomic Association, <http://www.ica.cc/ergonomics/> (Accessed on April, 2014)
- [2] Beith. B. (2014). Medical Device and Design Industry magazine, <http://www.devicelink.com/mddi/archive/99/06/009.html> (Accessed on May, 2014)
- [3] Sanders, M. S., Peay, J. M. (1988). "Human Factors in Mining", Report No. USBM IC 9182, United States Bureau of Mines, Washington, DC.
- [4] HSE, (2000). "Improving maintenance a guide to reducing human error" 1-61, ISBN 978 0 7176 1818 7
- [5] Dhillon B. (2008). "Mining Equipment Reliability, Maintainability, and Safety". Mining Equipment Reliability, Maintainability, and Safety (Springer Series in Reliability Engineering), ISBN-13: 978-1848002876
- [6] HSE, (2012). "Human factors: Maintenance error" <http://www.hse.gov.uk/humanfactors/topics/error.htm> (Accessed on April, 2014)
- [7] Warfield, J.W. (1974), "Developing interconnected matrices in structural modeling". IEEE Trans Syst Men Cybern 4(1):51-81

- [8] Sage, A.P. (1977), "Interpretive structural modeling: methodology for large-scale systems". McGraw-Hill, New York, 91–164
- [9] Singh, S., Kumar, R. and Kumar, U. (2014), "Modelling factors affecting human operator failure probability in railway maintenance tasks: an ISM-based analysis", *International Journal of System Assurance Engineering Management*, DOI 10.1007/s13198-014-0255-0
- [10] Attril, R. N., Devl and Sharma. (2013). "Interpretive structural modeling (ISM) approach: An overview", *Research Journal of Management Sciences*, , Vol. 2(2), 3-8,
- [11] Saxena, J. P., Sushil, and Vrat, P. (1990). "The impact of indirect relationships in classification of variables—A MICMAC analysis for energy conservation", *System Research*, 7(4), 245–253
- [12] Sharma, H. D., Gupta, A. D. and Sushil. (1995). "The objectives of waste management in India: A future inquiry", *Technological Forecasting and Social Change*, 48, 285–309.
- [13] Agarwal A., Shankar, R. and Tiwari, M.K. (2007). "Modeling agility of supply chain", *Industrial Marketing Management*, 36, 443–457
- [14] Mandal A and Deshmukh S.G. (1994), " Vendor Selection Using Interpretive Structural Modelling (ISM)", *International Journal of Operations & Production Management*, 14(6), 52-59,
- [15] Faisal M.N., Banwet D.K., Shankar, R. (2006), "Supply chain risk mitigation:Modeling the enablers", *Business Process Management Journal*, 12(4), 535-552
- [16] ISO9001, (2008). "International Standard Organization, Quality Management Systems Requirements".
- [17] Bridges,W. and Tew,R. (2010). "Human factors elements missing from process safety management (PSM)", Prepared for Presentation at American Institute of Chemical Engineers 2010 Spring Meeting and 6th Global Congress on Process Safety and the 44th Annual Loss Prevention Symposium, San Antonio, Texas, pp. 1-24
- [18] Guidance Note QGN14. (2008),"Effective Safety and Health Supervision",Queensland Government Department of Mines and Energy, pp. 1-37
- [19] HSE, (1992) "Workplace health, safety and welfare" ISBN 978 0 7176 0413 5
- [20] Employing People. (2014), www.gov.uk/browse/employing-people (Accessed on May, 2014)
- [21] Blink R. C. and Schreiberstein J.,(2007) "Fitness-for-duty evaluation: Walking the medical-legal tightrope". *Journal of Chemical Health and Safety*, vol. 14, pp. 9-13.
- [22] Dhillon B. (2007). "Human reliability and error in transportation systems ". (Springer Series in Reliability Engineering),ISBN13:9781846288111
- [23] Rooney J. J., L. N. V. Heuvel, and Lorenzo D. K.,(2002) "Reduce human error," *Quality progress*, vol. 35, pp. 27-36.
- [24] Energy Institue,(2008). "Guidance on investigating and analyzing human and organizational factors aspects of incidents and accidents", Published by Energy Institute.London, Registered charity Number 1097899, pp. 1-71.
- [25] Kumar S., Luthra S., and Haleem A. (2013)."Customer involvement in greening the supply chain: an interpretive structural modeling methodology," *Journal of Industrial Engineering International*, vol. 9, pp. 1-13.
- [26] Mason,S. "Improving maintenance by reducing human error ", Principal human factors consultant health, safety & engineering consultants limited (HSEC)

Asset Management of Aging Process Plant

John Andrews
University of Nottingham
NTEC, University Park
Nottingham, UK
+44 (0)115 84 68448

john.andrews@nottingham.ac.uk

Claudia Fecarotti
University of Nottingham
NTEC, University Park
Nottingham, UK
+44 (0)115 84 67220

claudia.fecarotti@nottingham.ac.uk

ABSTRACT

This paper reports a combined Petri net and Bayesian network approach to set the best asset management strategy. Petri nets have the ability to account for the detailed asset management strategy. This would include: inspection/testing, servicing, repair, replacement, renewal and enhancement. Inspection and testing includes the situation where the status of elements in the system is established using remote sensors. The results of the Petri net modelling are integrated within a Bayesian network to predict the likelihood of failure of system operation under different maintenance regimes. The integrated Petri/net and Bayesian network approach is demonstrated using a remote un-manned wellhead platform.

Keywords

Asset management, e-maintenance, remote sensors, Petri nets, Bayesian networks, fault diagnostics.

1. INTRODUCTION

Due to the cost of large-scale infrastructure construction many asset intensive industries operate their plant and systems beyond the originally intended design life. It is frequently more cost effective to use maintenance to keep the condition of the infrastructure in an acceptable state than to renew or replace it. The nuclear industry, the transport industries, the offshore oil and gas industry along with the water and power utilities provide examples of such situations. Controlling the asset state using maintenance requires an asset management strategy which is optimised to yield minimal whole life costs. Adequate asset management of aging plant is a critical requirement to prevent service disruption through breakdown but also essential for safety critical systems where failure can result in fatalities. E-maintenance can contribute to the efficiency of the process if sensor measurements indicating the state of the plant are taken and transmitted for analysis to indicate when interventions are required. The availability of this information can be incorporated into the modelling of the plant when the overall maintenance strategy is planned.

When assessing the adequacy of safety systems the risk of catastrophic failure is predicted and reviewed against target levels of performance. This process will involve identifying the frequency of occurrence of initiating events which start the incident sequence and the likelihood of the safety system(s) failing to respond. The analysis will take into account both the structure of the system (use of redundancy and diversity) along with the maintenance strategy employed to control the state of the

plant. Traditionally this analysis is performed using techniques such as integrated Event Tree [1] and Fault Tree Analysis [2-4]. The methods, whilst in widespread use, assume independence of the basic component failure events and the subsystems represented on the event tree. The maintenance options available govern the inspection/testing, servicing, repair, replacement and renewal. In addition there is, increasingly, the ability to remotely monitor the condition of some elements of the system. Opportunistic maintenance is also a possibility where work is carried out on components because the system is down due to other components requiring attention. Fault tree and event tree methods are not capable of modelling these factors in detail. An alternative method which has proved capable of modelling this sort of detail is the Petri net (PN) technique [5-7]. PNs constructed to predict the system performance based on the system structure, along with the component deterioration process and the maintenance strategy frequently feature characteristics whose solution requires the Monte Carlo technique [4]. It is therefore advantageous, in the interests of efficiency, to keep the size of such models to a minimum. This will effectively modularise the analysis process into separate, smaller, independent units. Bayesian Networks [8-10] are then capable of incorporating the results of the PN analysis to predict the system level response.

The approach is demonstrated in this paper by application to an unmanned wellhead platform used in the offshore oil and gas industry.

2. CASE STUDY – UNMANNED WELLHEAD PLATFORM

Remote wellhead platforms are used to exploit smaller oil and gas reserves where a full processing platform is not justified. These smaller unmanned installations house remote wellheads. The fluids from this platform are transferred, via pipelines, for processing at other platforms. Protection systems are incorporated onto the remote platform to prevent high pressure surges from the wells progressing to the processing equipment. If the surge is above the pressure rating of the equipment then there is a risk of hydrocarbon containment failure. In addition to the safety concerns this would also cause a major disruption to the production. An example safety system used to protect against the pressure surge, illustrated in Figure 1, is constructed of two subsystems. The first is the ESD (Emergency Shut-down) system. It features three sensors (S1-S3) which monitor the pipeline pressure. These send signals indicating the pressure to the controlling computer (COMP1). A trip is issued as soon as two

out of the three sensors indicate an excessive pressure in the line. To instigate the trip the computer controller will open the vent valves (VV1, VV2) which de-pressurises the pneumatic lines to the ESD valve actuators. Since the valves are of the air-to-open type this will cause both ESD valves, ESDV1 and ESDV2, to close. As long as one valve closes it will prevent the pressure surge damaging the process equipment. The second sub-system is the HIPS (High Integrity Protection System) which is a redundant version of the ESDV sub-system.

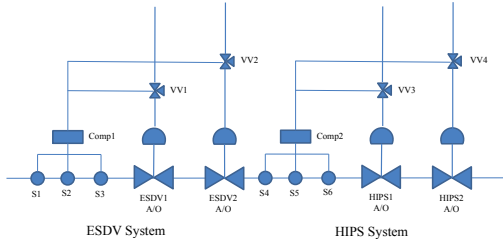


Figure 1. System Configuration

3. SYSTEM MAINTENANCE OPTIONS

The failures of concern for the safety system are, as usual for such systems, unrevealed when they occur and this failed condition is established by a testing regime carried out at set intervals when engineers visit the platform. The sensors, computer control unit, and the vent valves show no signs of a degraded state and are simply considered to work or not. For these components the maintenance processes will determine the test interval (θ_1) at which the component status is established.

The valves used on the ESD and HIPS sub-systems do however show signs of deteriorating performance in that the time it takes to close from their normal, fully open, position increases as the condition deteriorates and it gets closer to failure. From data, which has been established to relate the closure time to the remaining useful life, the valve can be tested by remotely (at intervals of θ_2) instigating an automated closure test where the duration of the valve closure is transmitted back to the maintenance centre. At the centre the required action can be determined. Clearly, the objective is to remove the valve from service prior to failure. In determining the maintenance strategy it has to be established at what condition (closure time) the valve is scheduled for **routine maintenance** (C_{ROUT}). If this is carried out too early in the valve life then it will discard the valve unnecessarily wasting some of the useful life remaining. If it is carried out too late in the valve life the risk of failure increases. Given that there is a substantial cost associated with sending a team out to perform maintenance on the remote platform, the condition monitoring brings about the potential for **opportunistic maintenance** where the valve will be replaced if work is to be performed on other components and its condition, whilst better than that at which replacement would usually be triggered, is less than some threshold (C_{OPP}).

Figure 2 shows the relationship between the deteriorating closure times for a valve and its remaining useful life. As can be seen, when functioning normally, the valve has a closure time of 10

secs. In this new condition it has a mean time to failure of 10 years. By the time the closure time has increased to 20 secs the mean time to failure has been established as 2 years. At 30 secs closure time the mean time to failure has become 1 year.

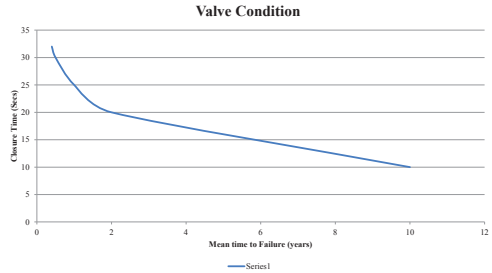


Figure 2. Valve Condition against Mean Time to Failure

4. SYSTEM AVAILABILITY MODEL

Traditionally for safety systems techniques such as Event Tree Analysis and Fault Tree Analysis are used to estimate the likelihood of system failure. This would assume that all of the components fail independently of each other. With the asset management strategy adopted for the remote platform the condition of the valves can be assessed remotely and this enables opportunistic maintenance to be carried out on the one valve dependent upon the condition of another. This introduces dependencies that require a more sophisticated modelling approach to determine their failure likelihood. A Petri net approach is used for this purpose. A Petri net, of similar structure, is constructed for the two valve systems (HIPS and ESD) which provides modularisation to the analysis. These sections of the system, whilst featuring dependencies between the two valve's failure probabilities, are independent of the rest of the system.

Since all other components fail independently, the commonly used equation [4] for unrevealed failures can be used to calculate their average probability of failure (Q_{AV}):

$$Q_{AV} = \lambda \left(\tau + \frac{\theta}{2} \right) \quad (1)$$

where λ is the failure rate, τ is the mean time to repair and θ is the inspection interval.

Having evaluated the Petri net models for the valve failure likelihoods these can be combined with the other components in the HIPS and ESD systems using a Bayesian Network. The structures of these two models are described in the following sections.

4.1 Petri Net Model

The valves can be modelled as two independent sections: the ESD valves and the HIPS valves. These components need to be modelled in order to consider opportunistic replacement since this introduces a maintenance dependency of one valve on the condition of another.

The ESD Sub-system Valves Model

The Petri net model for the ESD Valve sub-system is shown in Figures 3a and 3b. The model for the HIPS sub-system is identical in structure.

The PN model illustrated in Figure 3a shows the valve degradation process. Place P1 represents the valve in a state where it can be considered to be in good working order with no need of maintenance. This is defined by the closing time being around the 10 seconds expected. Should its condition deteriorate to a state where the closing time is considered to be slow (place P3) then we will instigate maintenance to replace the item prior to

its failure. If however the valve has not quite reached this state but its movement times are showing signs of deterioration (P2) it may be cost effective to carry out preventive maintenance on the valve should a team be sent to the platform to carry out work on the other valve (opportunistic maintenance). The definitions of the states P2 and P3 in terms of their closure times is a matter of selection when setting the maintenance strategy. Should the valve performance continue to degrade it will fail in one of three modes (all assumed to be equally likely). The first mode is a spurious failure where it fails resulting in closure or blockage to the line (Failed Closed). In this case the flow from the well will be halted, it will be immediately revealed, and attention required urgently.

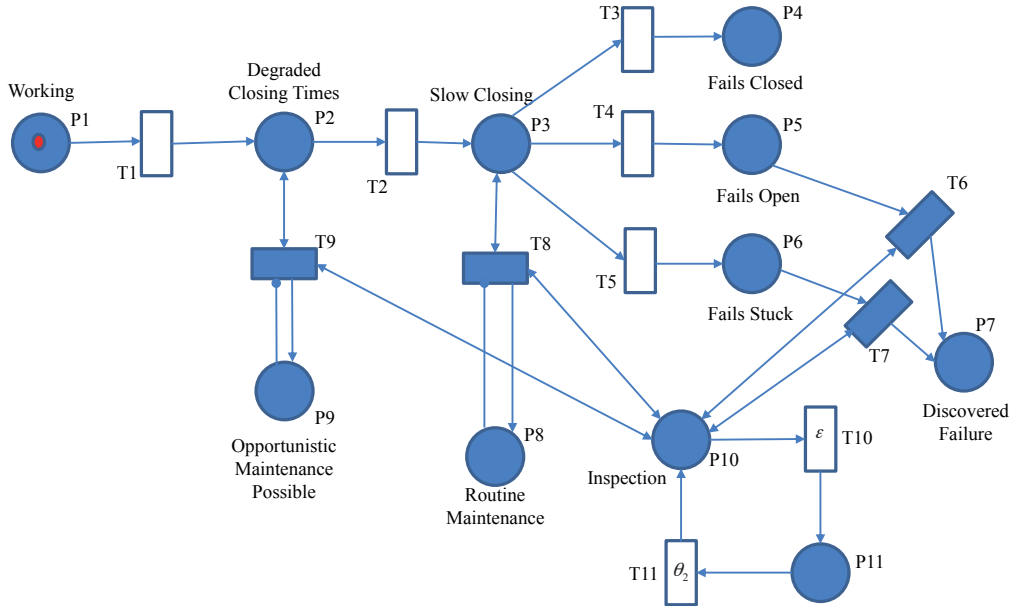


Figure 3a. Petri Net for the ESD Valves

On immediate occurrence of this failure mode a token is placed in P4 indicating emergency repair is required. The alternative, unrevealed, failure modes are: Failed Stuck and Failed Open. If the valve fails stuck then it will fail stuck in some intermediate position, which allows fluid to pass but cannot respond to a pressure surge. If it fails open then again it fails in a mode unable to respond to the pressure surge but in the fully open position. Both of these conditions will be identified when the remote tests are performed to reveal the closing time. Once revealed these two

states both result in a token being placed in P7 indicating a discovered valve failure. The testing process is represented by the loop of P11-T11-P10-T10- P11. When the token is on P10 it indicates that a test is taking place to reveal the performance condition of the valve. On revealing the condition of a degraded but not failed performance a token will be placed in P9 indicating that opportunistic maintenance is appropriate or P8 that routine maintenance is required for the valve. This instigates the appropriate scheduling of the preventive maintenance work.

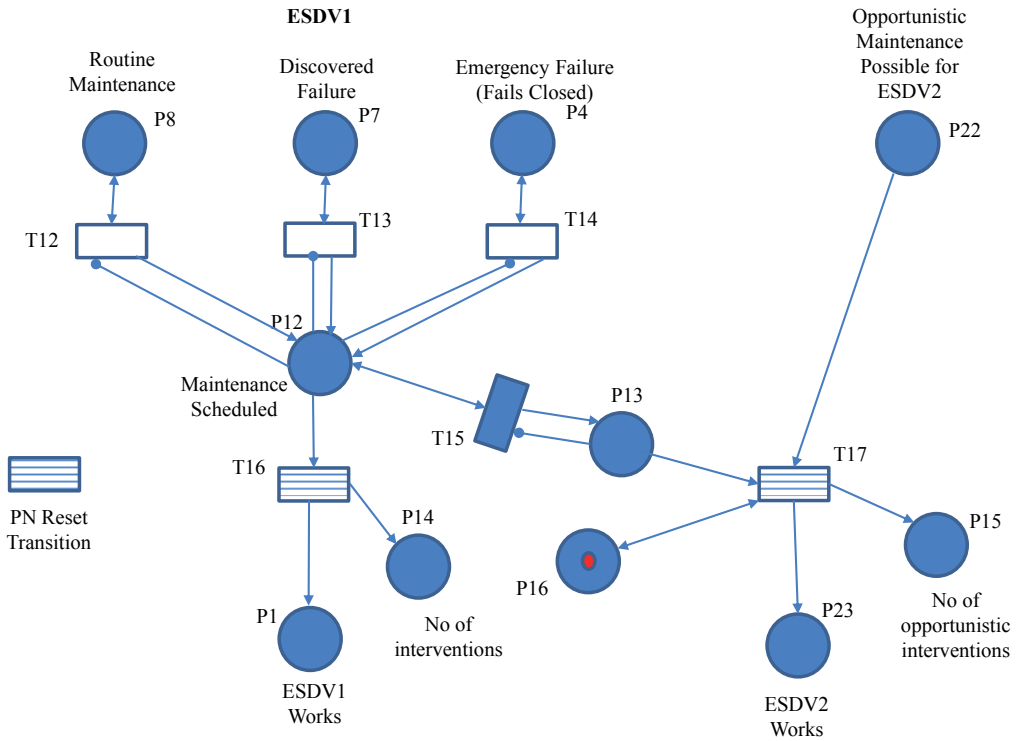


Figure 3b Petri Net for the ESD Valve Maintenance

Figure 3b indicates the PN for the work scheduling and maintenance, which restores the valve to the ‘good’ condition. States P9, P8 and P7 indicating ‘routine maintenance’, ‘discovered maintenance’ and ‘emergency maintenance’ respectively, cause appropriate work to be scheduled according to the urgency. This scheduling (making appropriate resources and transport available) has execution times appropriate for the seriousness of the condition and are set as part of the maintenance strategy definition. Note the reset transitions initialise parts of the PN when a repair has been instigated and a valve returned to the good state.

4.2 Maintenance Strategy Options

The number of options to be specified for the maintenance strategy are:

- [TIMan] Testing interval of the components (sensor, computer, vent valve) by maintenance personnel (θ_1): 3, 6 or 12 months
- [TIRemote] Testing interval for the sub-system activation closure times (θ_2): 3, 6 or 12 months

- [LOP] Closure times above which the valve can receive opportunistic maintenance (C_{OPP}): 20 or 25 seconds.
- [LR] Closure times above which the valve will be set for routine maintenance (C_{ROUT}): 30 or 32 seconds.
- [STRO] The scheduling time for routine maintenance ($Rout_{SHD}$): 2 or 6 months.
- [STRE] The scheduling time for revealed maintenance ($Revd_{SHD}$): 1 or 2 weeks.
- Enable opportunistic maintenance or not (OP): yes or no

The time to perform the valve repair is fixed at 1 day. The scheduling time for emergency failures will also be fixed at one day due to its priority. Failure rates and repair times for the other components in the sub-systems are provided in Table 1.

Table 1. Component Failure and Repair Data

Component type	Codes	Failure rate (per hour)	Mean time to repair (hours)
Vent Valve	VV1,VV2,VV3,VV4	1.0×10^{-5}	24.0
Computer Controller	COMP1, COMP2	1.0×10^{-6}	12.0
Pressure Sensors	S1-S6	1.0×10^{-4}	12.0

4.3 Bayesian Network Model

Analysing the PN models for the two valve systems and evaluating equation 1 for all other components will give the failure probabilities for each element in the system. The Bayesian Network (BN) Model combines these performance parameters for these components of the HIPS and the ESD sub-systems to give

the overall likelihood of an unrevealed safety system failure. The network, illustrated in Figure 4, has a clear hierarchical structure, which incorporates the PN model results for the sections with dependencies, the independent component failures and the selected asset maintenance policy.

At the top of the BN diagram is a node, which represents the system performance and considers two states, works of fails, in the event of a high pressure surge in the well. The system performance is dependent upon the performance of the two sub-systems (HIPS and ESDV), which are represented by nodes of the second level. Either of these systems has the ability to control process parameters in this case it is through the opening and closing of the valves and so the third level is the status of the sub-system active components i.e. the valves. Level four has nodes, which represents the functioning and failure modes of all of the components. This is linked to the nodes on the level above using a conditional probability table, which can be mapped in structure to represent the fault tree developing the causes of failure for the two valves [11].

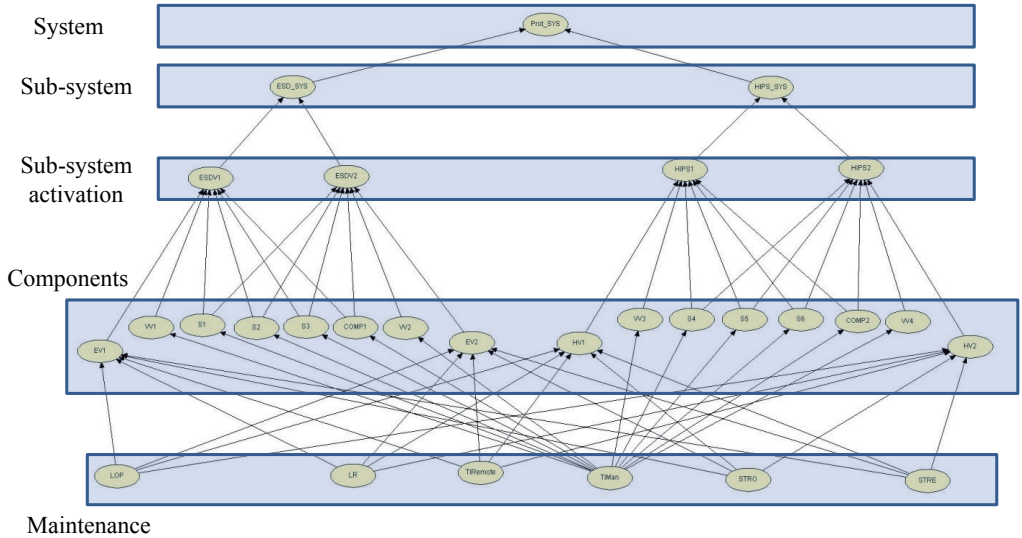


Figure 4. Bayesian Network Model

The final level is the level, which specifies the asset management strategy. There is one node for each of variables 1-6 listed in section 3.2. The final variable, allowing or not opportunistic maintenance effectively turns node LOP off in the network, which is reflected in place P16, (Fig 3b) in the Petri net. The Bayesian network shows the causality of changes in the system and which events influence the likelihood of occurrence of others. There are two basic event types those, which are dependent upon the status of other events and those which are root nodes (in this BN those which represent the asset management strategy options). For the

dependent nodes the BN will contain conditional probability tables (CPTs). An example of a CPT is shown in Table 2 for the sub-system layers ESD_SYS node which is dependent upon the state of nodes ESDV1 and

ESDV2. In this case the table is basically deterministic as if either ESDV works then so does the sub-system and so the probabilities entered are all 1 or 0. This will be the case for the CPTs for all nodes except those which appear on the Components layer of the BN. In this case the probability entries in the DCTs

will be evaluated from the PN, in the case of the valves, and equation 1 for all other nodes. For the independent nodes on the base level of the BN there will be a table indicating the likelihood of it being in any of its possible states. As this layer represents the options of the asset management variables then it will be used to select different options. A CPT for the root node representing the manual inspection options, TIMan, is illustrated in Table 3. In the table illustrated the inspection interval is set to 6 months.

Table 2. ESD_SYS Conditional Probability Table

ESDV2	Works		fails	
	works	Fails	works	5. fails
ESDV1	1	1	1	0
ESD SYS works	1	1	1	0
ESD SYS fails	0	0	0	1

Table 3. TIMan Conditional Probability Table

TIMan	Probability
3 months	0
6 months	1
1 year	0

5. RESULTS

That PN model for the 2 valves was analysed using software written specifically for the project. It has the ability to automatically rerun the PN for all options of the relevant asset management input nodes. These probabilities are then used to create the entries in the CPTs for the valves. The BN analysis was carried out using the HUGIN software tool. This enables the root node asset management options to be selected and predicts the likelihood that the system will be unable to respond to the pressure surge in the pipeline.

Petri net results

The main role for the PN analysis has been to generate probabilities for the CPT in the BN for the valves. However the results from these analyses are themselves very informative. For any place on the PN the number of times it is entered can be recorded, as can the duration of each residence time. As an example of the results which are provided by the PN consider two runs of the model. The first run is for the situation where no opportunistic maintenance takes place, the test interval is 6 months, the closure time which results in the request for routine maintenance is set to 30 seconds and the time to then schedule and routine and revealed maintenance is 2 months and 1 week respectively. When the PN model was run for 500 simulations of a 40 year system life the following results were obtained. For the second run opportunistic maintenance is allowed and the valve closure time which permits this is 20 secs.

The results of the PN analysis are provided in table 4 for ESDV1 and ESDV2 in each case. From the symmetry of the system structure the results for the two valves should be the same, the difference is due to the variability of a Monte Carlo simulation but are in close agreement. The results of the first two numerical

columns provide the average probability that valve is in the failed closed condition and the number of occurrences of this situation in 40 years. This will cause a disruption in production. As expected both of these figures reduce when opportunistic maintenance is allowed. Also reported in the table are the average probabilities that the valves are in the unrevealed failed state awaiting repair (discovered failure) and in a degraded state awaiting maintenance. The final two columns in the table indicate the number of times a repair is conducted on a valve which is not opportunistic and the number which are opportunistic. As can be seen the average total number of repairs is the same but the number of repairs which are opportunistic saves the maintenance team a helicopter trip to the remote platform.

Bayesian Network results

Populating the BN with the results of the PN runs enables the evaluation of any asset management strategy by selecting and setting the appropriate options in the root nodes and performing the quantification. For example, if the valve closing times that define the opportunistic and routine maintenance states are set to 20 and 30 seconds, the manual inspections performed every year with the automatic valve test carried out every 6 months, the schedule times for routine and revealed failure modes set to 2 months and 2 weeks respectively, an evaluation of the BN shows that the likelihood of the system failure in event of a surge is 0.1813. If we change the manual inspection to be performed every 3 months the system failure probability becomes 0.0014. In such a way the possibilities for the asset management parameters can be investigated.

6. CONCLUSIONS

An integrated Petri net and Bayesian network modelling approach has been developed for use in setting the system asset management strategy. It has been demonstrated by application to an unmanned wellhead platform. The method has the following advantages:

- The asset maintenance strategy can be investigated by using a Petri net model, which has the capabilities of incorporating the required level of detail.
- Since the Petri nets are solved using a Monte Carlo technique they are kept as small as possible by modularising the system into independent sub-units. This will result in an efficient analysis.
- The results from the independent units are combined using a Bayesian Network Approach in order to predict the whole system performance.

7. ACKNOWLEDGMENTS

John Andrews is the Royal Academy of Engineering and Network Rail Professor of Infrastructure Asset Management. He is also Director of The Lloyd's Register Foundation* Centre for Risk and Reliability Engineering at the University of Nottingham. He gratefully acknowledges the support of these organisations.

* The Lloyd's Register Foundation (LRF) supports the advancement of engineering-related education, and funds research and development that enhances safety of life at sea, on land and in the air.

Table. 4 Results from Petri Net Example Runs

		Valve failed Closed		Discovered Failure (Average Probability of P7)	Routine Maintenance (Average Probability of P8)	Opportunistic Maintenance (Average Probability of P9)	Average Number of Conventional Repairs	Average Number of Opportunistic Maintenance
		Average probability	Average Number					
No opp	ESDV1	5.65e-5	1.856	0.0064	0.00578	0.0	3.876	0
	ESDV2	6.18e-5	1.812	0.0065	0.00585	0.0	3.770	0
opp	ESDV1	5.55e-5	1.554	0.0065	0.00570	0.0105	3.502	0.288
	ESDV2	5.13e-5	1.592	0.0064	0.00533	0.0112	3.526	0.286

8. REFERENCES

- [1] Andrews, J.D. and Dunnett, S.J., (2000). "Event Tree Analysis Using Binary Decision Diagrams", *IEEE Transactions on Reliability*, 49(2), pp. 230-238. ISSN 0018-529.
- [2] Rauzy, A., (1993) "New Algorithms for Fault Tree Analysis", *Reliab. Engng. Syst. Safety*, 40, pp. 203 - 211.
- [3] W.E. Vesely, (1970) "A Time Dependent Methodology for Fault Tree Evaluation", *Nuclear Design and Engineering*, 13, pp. 337-360.
- [4] Andrews, J.D. and Moss, T.R., (2002), "Reliability and Risk Assessment (2nd edition)". *Professional Engineering Publishing*, 540 pp, ISBN 1-86058-290-7.
- [5] Jensen, K., (1997), "Coloured Petri Nets", *Springer Verlag*.
- [6] Schneeweiss, W.G., (1999), "Petri Nets for Reliability Modelling", *LiLoLe Verlag Publishing Company Limited*.
- [7] Audley, M. and Andrews, J.D., (2013), "The Effects of Tamping on Railway Track Geometry Degradation", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 227(4), pp 376 – 391.
- [8] Jensen, Finn V; Nielsen, Thomas D., (2007) "Bayesian Networks and Decision Graphs". *Information Science and Statistics series (2nd ed)*, New York, Springer-Verlag.
- [9] Doguc, O. and Ramirez-Marquez, J.E., (2009), "A Generic Method for Estimating System Reliability Using Bayesian Networks", *Reliability Engineering & System Safety*, 94(2), pp. 542–550.
- [10] Langseth., H. and Portinale, L., (2007) "Bayesian Networks in Reliability", *Reliability Engineering & System Safety*, 94,(2) pp. 92-108.
- [11] Bobbio, A., Portinale, L., Minichino, M. and Ciancamerla, E.,(2001), "Improving the Analysis of Dependable Systems by Mapping Fault Trees into Bayesian Networks", *Reliability, Reliability Engineering & System Safety*, 71, pp. 249-260.

Issues and challenges for condition assessment: A case study in mining

Esi Maan Nunoo

Division of Operation, Maintenance
and Acoustics

Luleå University of Technology

Luleå, SE-971 87, Sweden

+46 920 493427

esi.nunoo@ltu.se

Phillip Tretten

Division of Operation, Maintenance
and Acoustics

Luleå University of Technology

Luleå, SE-971 87, Sweden

+46 920 492855

phillip.tretten@ltu.se

Aditya Parida

Division of Operation, Maintenance
and Acoustics

Luleå University of Technology

Luleå, SE-971 87, Sweden

+46 920 491437

aditya.parida@ltu.se

ABSTRACT

Organisations need to continuously review their costs to be profitable. Thus, they need an understanding of the condition of their facilities, their equipment and processes, so that they can make good operation and maintenance decisions. The mining industry in Europe relies on complex technical systems for production. Error conditions in equipment will result in a failed system, having drastic effects on production. These failures can impact business performance, resulting in increased costs due to reduced availability, reduced productivity, product quality and safety. Condition assessments of failure data can be helpful in predicting the health condition of equipment in the near future, thus helping to make efficient planned stops. The purpose of this study is to identify the issues and challenges for condition assessment in the mining industry. A case study was conducted using relevant literature, interviews with operation and maintenance personnel and the production process was studied. The findings showed that the physical conditions of system components are not always known nor is the useful life of the components known. This is due to the fact that in spite of the large volumes of data collected on critical equipment, there is little analysis done on the data to be able to predict future failure patterns or failure trends. Maintenance on equipment is often of corrective nature. Using eMaintenance for condition assessment will help to make failure more predictable thus optimizing operation and maintenance decisions.

Keywords

Maintenance, eMaintenance, Condition Monitoring, Condition Assessment, Key Performance Indicators

1. INTRODUCTION

Global competition requires that organisations are able to have uninterrupted plant availability and reliability in order to maintain a competitive advantage. The past 50 years has seen maintenance responding to the changes in user expectations and an evolution from corrective maintenance to predictive maintenance. This growth is attributed to the connection between maintenance and product quality, coupled with the increasing pressure to achieve high plant availability, reliability and cost effectiveness [1]. The consequences associated with plant failure not only affects the total level of productivity but can also affect product quality, safety, the work environment and the sustainability of the organisation [2].

Organisations need to continuously review their asset performance to be competitive and profitable. Thus, they need to have an understanding of the condition of their facilities, engineering assets and processes, so that they can make good operation and maintenance decisions. Making the right maintenance decision not only improves asset performance, but also reduces the cost of maintenance, the cost of unplanned stops and the cost of total production. The wrong choices on the other hand could create new problems while escalating existing problems [1]. Maintenance is important because of its ability to retain and improve system availability and reliability, performance and safety, as well as, the quality of products and services [3].

The mining industry in Europe relies on complex technical systems for production. As the physical assets grow in complexity, so does their cost of ownership and operation. To secure the maximum return on investment, plants must work efficiently for as long as the owners want them to [1]. Remaining a leader in the mining industry requires optimal production levels, as well as, high quality products.

This can only be achieved when plants are available and reliable. Unfortunately, plants will not always be available or reliable if decisions are made without taking into consideration condition monitoring data, condition assessment and analysis of the data gathered. As plants get older or are operated incorrectly they may experience frequent failures. These failures may be as a result of wear and tear and may require preventive and predictive maintenance in order to control unplanned stops. Monitoring the condition of equipment and machinery makes it possible to predict potential failures, thereby preventing or reducing unplanned stops. Being able to assess the health status of the equipment means being able to plan and schedule for possible failures so as to prevent plants from completely breaking down.

The cost of maintenance can immensely affect the total cost of production of a product. Research indicate that maintenance costs for the iron and steel industry, the pulp and paper industry, and other heavy industries represent up to 60 percent of the total production costs [4]. Though it is almost impossible to get accurate figures on the cost of maintenance in the mining industry, Campbell [5] suggests that in a highly mechanized mine, maintenance can cost as high as 40% to 60% of the operating cost. Murthy et al, [6] also suggest that the annual cost of corrective and predictive maintenance as a fraction of the total operating

budget can be as high as 40% to 50% which translates into \$500 million per year for a big mining company. According to Cross [7], UK's maintenance spending in the manufacturing industry ranges from 12% to 23% of the total factory operating cost while Altmannshoffer [8] affirm that Europe spends about 1500 billion euros yearly as part of their maintenance expenditure. Though direct cost of maintenance is high, the indirect cost of maintenance is much higher. Indirect costs could arise as a result of a loss in production, a delay in delivery or customer dissatisfaction, which could lead to a loss of goodwill and customers [6]. The slightest reduction of maintenance cost could reduce operating costs and improve product quality, which can eventually contribute to an organisation's profit [9].

Considering the effects of direct and indirect cost of maintenance on the total cost of production, it is prudent to make use of maintenance techniques that can in the long run reduce the cost of production and retain or improve product quality while preventing and/or reducing unplanned stops. In order to prevent unplanned stops, there is the need for the mining industry to be able to monitor and control failures by way of implementing condition monitoring and assessment techniques

The objective of this paper is to study the issues and challenges for condition assessment on critical equipment in a mining industry.

The structure of the article is as follows: section 1 gives an introduction to the problem. Section 2 provides literature on what has already been done with regards to condition assessment, asset management, issues and challenges for condition assessment and information management and effective condition assessment. Section 3 presents the case study while section 4 discusses the findings of the study. Section 5 concludes the study.

2. LITERATURE REVIEW

The growth in research in maintenance is due to the advances in the understanding of the physics of failure, technologies to monitor and assess the physical health status of systems, in computers to store and process large amounts of relevant data and in the tools and techniques needed to build models to determine optimal maintenance strategies [10].

Corrective maintenance, though important and necessary, is not always the best option as it could lead to unplanned stops and a great loss in output produced, where unplanned stops are prolonged. Corrective maintenance costs as argued by [4] could cost three times more than the same repairs made on a scheduled basis. Downtime affects productivity by reducing output, increasing operating costs and interfering with customer service [1] as plants may break down during peak periods. Condition monitoring and assessment makes it possible to predict future failures, thereby allowing for planned stops during which such failures could be corrected. This helps to retain production at normal levels and at the same time maintain the plants.

2.1 Condition Assessment

Condition assessment is the process of measuring the physical condition of system components, using specific, clearly defined observable and measurable indicators [11]. Condition assessment can also be defined as the task of evaluating the current condition of a component, a system or a plant based on observed, recorded and/or reported characteristics. This maintenance assessment can

be achieved through the use of technologies that are able to determine equipment condition using historical failure records or rate of failure to predict failure before it happens. In essence condition assessment can be achieved through condition monitoring methods such as analysis, process monitoring, performance monitoring, functional testing and inspection [12].

Condition assessment is important because it can assist with developing and understanding the extent of problems associated with a system. It can also be used to better target the investment requirement needed to maintain the optimal quality of products and services. Condition monitoring systems help decrease the operational risk, enhance the performance of a system or a plant and eventually increase the speed of production. This ultimately leads to a reduction in the cost of maintaining the plant, which in the long run leads to a decrease in the cost of production and thus a reduction in the unit cost of the product.

Condition monitoring is a maintenance philosophy wherein equipment repairs and replacement decisions are made based on the current and projected future health of the equipment [13]. The data that is stored as a result of the system being monitored can be used for further analysis in order to improve the system as per Figure 1 below.

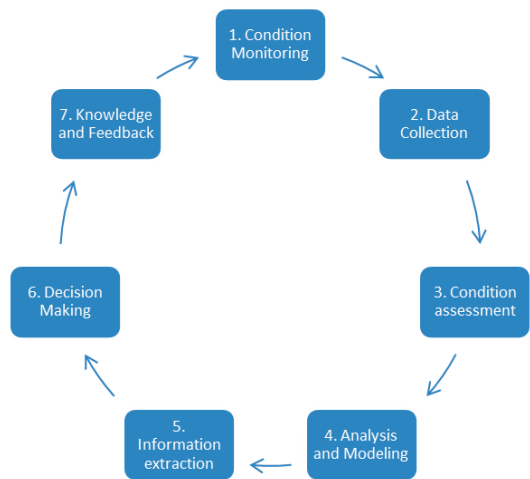


Figure 1. The Condition Monitoring Process

It is not only important to collect health status data of the assets, the data must be analyzed to generate knowledge that can help to take proactive steps toward increased reliability and availability of the assets. Information obtained in condition monitoring and assessment enables the determination of the remaining useful life (RUL) of assets. Knowledge of the RUL of assets enables efficient use of resources and assists in maintenance planning activities and decision making, spare parts provision and operational performance [14].

To improve availability, reliability, and to cut the cost of maintenance due to the level of mechanization, organisations are incorporating eMaintenance processes to their already existing traditional maintenance processes. eMaintenance is a multidisciplinary domain based on traditional maintenance processes and Information and Communication Technologies (ICT) ensuring that eMaintenance services are aligned with needs and business objectives of both customer and supplier and are inherent constituents in the whole system life cycle [15]. ICT allows the merging of existing maintenance processes with web services and other modern e-collaboration tools that can help foster the effectiveness and efficiency of the maintenance process. The development of computer technologies and artificial intelligence techniques makes it possible to implement condition monitoring more effectively, reliably, and intelligently [16].

To remain the market leader of mining products, different maintenance strategies are required. In addition to corrective maintenance, proactive maintenance is also required. Computerized predictive maintenance can help make the process of fault detection more effective. The right mixture of corrective maintenance and computer based predictive maintenance can go a long way to help an organisation increase the reliability, and availability of plants thus increasing the speed of production.

2.2 Effective Asset Management

Condition monitoring and assessment can lead to effective asset management as it provides the necessary information logistics needed to manage the equipment in use. Asset management is defined by [17] as a strategic, integrated set of comprehensive processes; financial, management, engineering, operation and maintenance to gain maximum lifetime effectiveness, utilization and return on investment from physical assets. Effective asset management helps improve reliability, utilization and availability thus improving operational effectiveness, and customer satisfaction, while reducing operating costs [18].

Knowledge of the health of plants and equipment help decision makers and resource planners make informed decisions, set prioritized objective and rational maintenance activities that is aligned with the needs and business objectives of both the customer and the supplier. Effective asset management extends equipment life and thus improves and optimizes business profitability [18].

Proactive maintenance through condition monitoring and assessment creates an opportunity for improvement by gradually extending equipment Mean Time between Failures (MTBF) due to improvement in the precision with which maintenance is performed [19].

2.3 e-Maintenance for real time information management

Information management deals with the collection and controlling of information from different sources and the dissemination of this information gathered to several audiences who need it.

In condition monitoring an asset, several characteristics of that asset is recorded for possible maintenance purposes. For condition assessment to be beneficial there is the need not only to acquire maintenance data but to synthesize and analyze this data to produce meaningful and timely information that can be used to make strategic, tactical, and operational decisions.

Data is only useful if it can be presented as qualitative and time sensitive information [20]. The application of ICT in condition monitoring and assessment could help facilitate real-time information management to the right audience. ICT based condition monitoring and assessment can also help to facilitate prognostics and decision support thus leading to the right type of maintenance activity being selected [21].

The use of ICT can help provide a means by which maintenance data from different sources irrespective of location and geography could be brought together where needed. This is true especially for a mining company that has several mines scattered across the country with several offices. eMaintenance solutions through web services can also provide a means through which manufacturers of plants can provide common and up-to-date FAQ's on common fault resolution to users of their equipment. eMaintenance solutions will facilitate maintenance decision-making by providing effective and efficient maintenance information logistic [22].

2.4 Issues and Challenges for Condition Assessment

Several eMaintenance techniques have emerged in the last decade. No matter the choice of technique, users expect some sort of support for e-operation. This support could be through remote diagnostics and asset management, simulation for optimisation and decision making under an e-business scenario for an organisation [22].

Voluminous information flow and system complexity can pose some challenges when using eMaintenance solutions [23]. Maintenance-related information is often hidden in vast data, stored for other purposes, at different places, in different formats, and generated throughout the entire life cycle of equipment [24]. The data gathered may need to be integrated to provide useful information. What data to store is very important as this could affect the quality of analysis and eventually decision making [22]. On the other hand, making any useful analysis from the data gathered where the application does not have support for analysis requires the help of experienced personnel [25]. Without experienced personnel, such condition monitoring programme does not yield the maximum benefits [21].

Another challenge of using eMaintenance in condition assessment is that managers do not always have time or personnel to process and analyze condition monitoring data as the workforce is dedicated entirely to dealing with breakdowns that occur in the shop floor [26].

Biswal and Parida [27] report that many companies still use their Enterprise Asset Management Systems (EAMS) or Computerized Maintenance Management System (CMMS) solutions as static data and reporting systems. They are not able to harness the real-time data management and alarming systems which are integrated into the EAM or CMMS package. For this reason they do not benefit from a real-time and dynamic knowledge driven EAM/CMMS solution.

Most often than not, proactive maintenance through high technology does not yield much measurable benefits because of failure to make the necessary changes in the work place that would permit maximum utilization of these predictive tools [4].

Biswal and Parida [27] report Reliability Engineering online magazine as saying that the major reason why most of the predictive maintenance programs fail is because the goals of the program is not well defined or understood: these goals are neither aligned to the exact needs for performing predictive maintenance in the organisation nor are they aligned with the needs and business objectives of both customer and supplier.

Fernandez, Labib, Walmsley, & Petty [26] report that though a computer system capable of generating predictive maintenance schedules may be in use, its operation could be complicated and time consuming for storing maintenance data. The system may also have reports modules that are so rigid that users are unable to obtain the information they need [26].

eMaintenance solutions tailored to the need of the user could create opportunities for measuring the physical condition of system components, using defined, observable and measurable indicators on condition monitored data.

3. CASE STUDY

This case study was undertaken on a critical plant in a European mining company. Besides, using relevant literature, interviews with operation and maintenance personnel, automation unit personnel, production unit personnel and personnel from the IT section were undertaken. Literature related to Maintenance, eMaintenance, condition monitoring, and condition assessment were studied.

The company currently has a Maintenance System (CMMS) for creating their work orders. The CMMS is used to manage information about the organization’s maintenance operations. They also use another computerized system for capturing the process control data and another for monitoring faults and following up failures in production.

No. of Interviewees	Role	Experience
Interviewee 1	Maintenance Manager	3 years
Interviewee 2	Maintenance Engineer	16 years
Interviewee 3	Productions Manager	12 years
Interviewee 4	Process Engineer	4 years
Interviewee 5	Automations Engineer	14 years
Interviewee 6	Personnel from OVMC	11 years
Interviewee 7	Management Conductor	20 years
Interviewee 8	Sections Manager(IT)	10 years
Interviewees 9 (2 People)	Maintenance Engineer	3 years
	Manager	1 year

Table 1. Interview Schedule

The company has outsourced vibration condition monitoring and diagnostics of the critical equipment to an Online Vibration Measurement Company (OVMC). All data gathered during CM by this company is kept with them. The mining company has no

way of accessing the data because they do not want to be overwhelmed with it. They are furnished with reports as and when they require it by contacting personnel at the online vibration measurement company.

The goal of this case study was to gain a better understanding of the issues and challenges for condition assessment in a mining industry. In total, ten people were interviewed; four personnel from the operation and maintenance section, two personnel from the production unit, one personnel from the automation unit, one personnel from the process unit, one personnel from OVMC and one personnel from the IT section. Table 1. presents the interview schedule indicating the role of the interviewee and the number of years of experience of the interviewee

The interviewees were experienced and had a good understanding of their working environment. The interviews were semi-structured with open-ended questions lasting on an average of about 43 minutes each. Questions used for the interview were focused on the maintenance process in the organization, condition monitoring and assessment and the issues and challenges faced in condition assessment for maintenance activities. All interviews were tape recorded and transcribed later in order to make data analysis easier and capture all relevant details given by the interviewees.

In order to see the failure prone areas on the equipment, a sample historical work order data of a period of four years was collected. This data was used to draw the Pareto chart in Figure 2.

3.1 Maintenance Issues on Critical Equipment

There is a high demand for the products that this mining company produces. Unplanned stops mean a loss of money amounting to loss of tons of products that could have been produced. The direct cost of maintenance in itself is not costly for this mining industry. What is costly to them is the indirect cost of failure especially in the critical equipment. The influence of unavailability on factors such as output produced, product quality, customer service, safety, the environment and operating cost makes its cost dominant in the total mining cost.

Below are some of the issues raised during the interviews about critical equipment.

1. The age of the equipment is not the only determinant of failure. Factors such as how it is used, whether it is overused or operated in the right way or is maintained with the parts recommended by the manufacturer can also cause it to fail.
2. Knowledge of the health status of the equipment helps the maintenance team to plan for maintenance stops thus improving the overall performance of the plant. If a fault on the equipment will not hold until the next planned stop, then an early planned stop could be scheduled to resolve the issue. Planned stops will not be possible if there is no knowledge of the health status of the equipment.
3. Condition monitoring devices presently monitor one condition at a time. Computerizing condition monitoring and assessment would help to analyze several data gathered by the different sensors. This can help predict possible failure and their effects on the plant as a whole. Apart from vibration monitoring, all other monitoring is physical inspection based. Data collected on such inspections are primarily paper based. There is more

emphasis on corrective maintenance than predictive maintenance. There is the need to increase other forms of automatic and computer based monitoring and assessments in order to improve on predictive maintenance and reduce unplanned stops.

4. Failure records enable the determination of the life span of equipment thereby making it possible to put in place preventive measures to reduce unplanned stops.
5. The complexity of a machine can reduce its predictability especially when there is no computerized system to help track the smallest change in the way the machine operates. Lack of computerized analysis of the health status of complex equipment can lead to an increase in random failures due to unpredictability.
6. It is hard to track changes to machinery when the majority of monitoring is paper based inspections. It is possible to miss important information that could help determine the potential-to-functional failure interval (P-F interval) for machines. Old plants get changed in parts of them. It is important that technicians are included in projects that involve rebuilding parts of these plants; this way they would not miss delivering this information to personnel who work with them.
7. Incorrect data can be entered into the CMMS because Maintenance Technicians after writing the faults identified on a sheet hand them over to the maintenance planners who later enter these faults into the CMMS. As this data changes hands, some mistakes can be made while entering them into the CMMS.
8. Some problems are only found during maintenance stops. If these were not found during the planned stops, then they could result in big problems when a new production year starts.
9. It is important for Maintenance Technicians on inspections to have access to the CMMS and be able to verify the status of work orders during inspection. This will prevent duplication of fault reports while on inspection and ensure that they are able to make entries to the CMMS while on inspection and print out important information from the system during inspections.
10. Using human senses to predict failure has its own draw backs. There is the need to have other in-house computerized analysis of data gathered. This will help compliment the knowledge generated out of the inspections that are done. Below are some of the issues and challenges of the involvement of humans in condition monitoring on equipment:
 - For the human to notice any failure in a system, that system would have advanced in deterioration.
 - The human noticing failure is subject to their experience and competence. Whereas an experienced person can see multiple faults with particular equipment in use, an inexperienced person may see none or just a few. Not seeing what could be wrong with the equipment could hinder seeing what is likely to happen as a result of not knowing what is happening in the present.
 - What to do with the work orders in the CMMS created by the Maintenance Technicians can also be problematic. Some personnel don't describe exactly what has happened and how to fix it whereas some also don't write clearly. If a problem is not well understood it cannot be fixed. Inconsistent and unclear writing, not understanding what has been reported would require extra time to confirm what is broken, where, and how it is broken or which parts are needed before any resolution can take place.
- The human seeing faults on time during inspections are also dependent on how well they know the equipment. This means that if they are new to the equipment they may never be able to tell what failures are eminent. A computerized monitoring and analysis system on the other hand can help even an inexperienced person to find faults thus increasing decision making and fault resolution thereby optimizing personnel use in other areas which need their attention more.
- Not having enough personnel to do inspection could also hinder fault detection. Inspection lists that are either not correct or too big such that the Maintenance Technicians cannot get time to go to all the inspection points as a result of insufficient Maintenance Technicians per allocation.
- Automation does not necessarily replace humans. People must still check to see if these automations are up and running every now and then. There are over 1000's of sensors and not many people to check how they are on a regular basis. Though the sensors must be checked at least once every year, maintenance personnel are not able to do this. In the case of this mining company, some of the sensors do not work properly and so the data they generate cannot be trusted. Unless the faulty sensors are changed and verified, the data they generate cannot be depended on. There is the need for more technicians so that inspections of the sensors can be catered for. With the existing number of technicians and their list of activities, it is not possible to check all the sensors.
- Separation of job roles or responsibilities could also affect output and decision making. The one who sees to availability is different from the one who produces the product. If the operator isn't experienced, it affects production.
11. Operation and maintenance of critical equipment depends on the experience of the personnel involved. Wrong operation and inadequate maintenance could lead to an increase in the number of unplanned stops. This is due to the fact that a small fault that is not resolved can cause a ripple effect in the whole system, giving way to new problems while worsening the old problem.
12. There is no way of creating customized reports that suits the needs of the organization when needed by personnel of the company. All reports are requested for by personnel from OVMC. This slows down decision making processes as personnel have to wait until they have been furnished with a report to know the extent of damage caused by faults on equipment.

3.2 Causes of Failure of Critical Equipment

As per the information gathered through the interviews, possible failure in the equipment could be due to:

1. Formation of slag in the equipment
2. Bearing faults
3. Poor quality of lubricants
4. Poor maintenance activities. The wrong choice of a maintenance technique can create new problems while increasing and sometimes complicating existing problems.
5. Inability to monitor the hot parts of equipment.
6. Inadequate knowledge of personnel.
7. Insufficient skill and training of the operators and technicians to maintain the equipment in use.
8. Wear and tear on equipment that are missed by the inspection personnel.
9. Uneven speeds of operation of the plants due to production quantity and quality issues.
10. Problem of running the plant with a mixture of new and old parts creating inconsistent speed.

All components of the critical equipment do not have the same level of criticality to production. Currently equipment is classified as A, B, C or D depending on the importance of the equipment to production. Equipment which fall in class A are considered the most important to production, as failure in any of these would mean a stop in production or a shutdown. Equipment in class B are also important but not as those in class A. Class C to D equipment are those whose failure will not cause production to halt.

In order not to increase the cost of condition monitoring and the quantity of data gathered there is the need to identify the fault prone areas of the critical equipment. Selecting the most critical equipment to production will speed up decision making processes as well as control the cost of condition monitoring to the organization. This will also reduce the quantity of data to be collected. As per the company policy only equipment in class A are monitored.

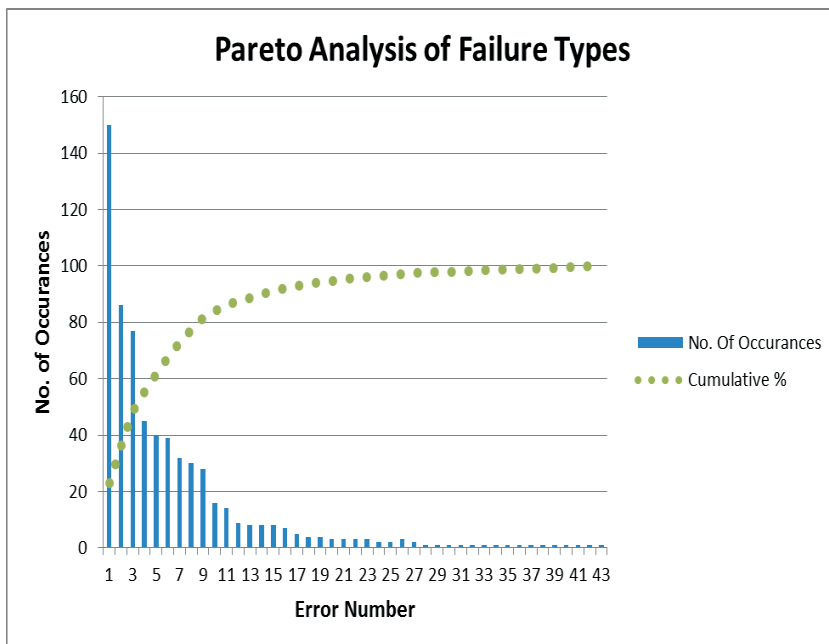


Figure 2. Pareto Chart of Faults on a critical equipment

For the condition assessment and monitoring programme to yield the best results, there must be a way to identify which equipment or which components of the equipment in use have the highest failure rates and their consequence on production. A typical way of doing this is the 80/20 rule. The 80/20 rule states that in anything a few (20%) are vital while many (80%) are trivial. From Figure 2, it can be seen that about 20 percent of the faults recorded are responsible for 80 percent of the downtimes. This means that if the 20% failure can be controlled, then unplanned stops can also be controlled. Faults registered under others, flat injury and scrappers are the most recorded of all the faults. Though the chart above does not show the cause of these faults, it helps to identify which fault may need attention in order to reduce unplanned stops. This can allow the maintenance personnel and management to focus on these faults since they seem to have the most occurrences out of the lot.

The question then is how do we control these 20% faults? One solution would be to monitor and detect these faults ahead of time in order to prevent their occurrence. Which condition monitoring techniques should be used then? Using paper based inspections, it would be difficult to track less change. A computerized condition monitoring and assessment programme would be able to analyze the data gathered more effectively and efficiently thus speeding up decision making processes.

There is currently an eMaintenance system doing vibration analysis on some of the equipment. The challenge with this current level of condition monitoring and assessment is that decision making is dependent solely on the availability of personnel from OVMC. The data gathered through vibration monitoring is not readily available and cannot be analyzed by the mining company.

3.3 Future Expectations

The expectation of the operation and maintenance personnel is to have:

1. A computerised Condition Monitoring and Assessment system with efficient Maintenance Information Logistic support. This will help to increase availability and speed of production. It will also reduce unplanned stops and improve decision making in the area of maintenance and production.
2. A computerized Condition Monitoring and Assessment solution will ensure that jobs planned are the jobs that need to be planned. Planning maintenance jobs ahead of time helps the maintenance personnel to choose the best approach for fault prevention and resolution.
3. There is the need for other condition monitoring programmes other than vibration as is currently being done. As production increases, insufficient monitoring and assessment could lead to an increase in the unplanned stops. Other condition monitoring techniques could help the maintenance personnel to see slugs early enough and remove them before they break off and interrupt with production.
4. Hand held computerized devices to replace the inspection sheets which do nothing but record faults. Sheets of paper cannot be used for much analysis while on inspection. There is the need for hand held devices that can help personnel to see what to check for, what has been checked for and reported and also allows for the creation of work orders

while on inspection. This will solve the problem of duplication of fault records, and help maintenance technicians and operators see the status of a work order thus knowing what more needs to be done for an existing work order.

5. An interface on OVMC's website where reports can be generated by the mining company when needed. This will improve decision making processes with regards to maintenance.
6. Improvement on how machines with slow revolutions per minute (RPM) are monitored as it is difficult to do this at the moment.

4. DISCUSSION

The cost of unplanned stops by far supersedes the direct cost of maintenance as it includes the cost of indirect cost factors such as product quality, customer dissatisfaction, safety, the extra operating cost and the loss of production. It is difficult and sometimes impossible to predict failures or to prevent unplanned stops when there is no historic data of the health of the equipment under review. It is also of no use when the quality of the data generated cannot be trusted because some parameters are missing or cannot be understood.

Condition assessment has numerous advantages. One of which is the extension of the MTBF due to improvement in the precision with which maintenance is performed. Condition monitoring and assessment ensures that jobs planned are the jobs that need to be planned within a time frame such that there will not be unplanned stops. Planning maintenance jobs ahead of time helps maintenance personnel to choose the best approach for fault prevention and resolution in order to prolong and increase availability of the plants and equipment.

Computerized and fully automated condition monitoring systems can help achieve better analysis than the manual ones like inspections. With a computerized system, little changes to the health of the equipment that would otherwise have been invisible can be noticed. Additionally, a computerized monitoring system can be easy for the less experienced personnel to work well with without much knowledge of the analysis involved.

The case study shows that the present condition assessment programme currently in use is not yielding its full benefits. Different systems are used to gather different data for different purposes. Consolidating these data to produce information for decision making can sometimes be tedious and daunting. Integration of process and condition monitoring data is important in timely decision making [22].

Data collected by OVMC is solely managed by them. The organisation has no control of this data and is furnished with fault reports as and when they need based on how soon personnel from OVMC can make this report available to them. The use of Web Technologies could provide information from integrated data analysis in real time to all stakeholders for decision making even without the use of a centralised database [21]. Integration through web services on OVMC's website could also provide a means for the mining company to be able to retrieve the reports they need as and when they need it without relying on personnel from this company to generate the reports for them.

Fault reports recorded on sheets of paper during inspection could lead to erroneous records entered into the CMMS by other personnel who depend on these sheets to create work orders. The use of sheets of paper for inspection could also encourage personnel who create work orders to leave fields empty due to unclear hand writing on the sheets used for inspection. Maintenance planners cannot always read what has been written and so may not be able to capture exactly what has been reported. Inadequate descriptions of the fault and what need to be done to fix them or what materials are needed to be able to fix these faults could prolong downtimes.

By using sheets of paper for inspections, it is not possible to know which faults have been reported, which have been worked on and which are in the process of getting fixed. Duplicate reports could be made on the sheets that personnel use while on inspections. Duplicate reports would require extra time for the maintenance planners to track as they would have to manually find the duplicates in order not to create more than a work order for the same fault reported.

The use of hand held devices could facilitate some level of automation during inspection thus eliminating the traditional paper based fault reports made during inspections. Applications could also be developed on the windows phone for personnel since all personnel own windows based smart phone for work. This would enable technicians have easy access to records from all the computerized systems needed and further help technicians who require the details of the work order in order to resolve the faults reported work more efficiently.

The current use of sensors to monitor vibration only or temperature only could be replaced by sensors that can monitor multiple conditions such as vibration and temperature. This could help minimize the total number of sensors in use. This way, monitoring the health of the sensors would be less stressful compared to thousands of sensors in use at the moment. The number of conditions to monitor should not be too many as the more conditions the sensor monitors might increase the sensor size making it bulky.

5. CONCLUSIONS

In this paper, the issues and challenges for condition assessment in a mining industry are discussed. It is not enough to monitor plants and equipment and not be able to effectively utilise the data generated for analytical assessments. Without analysis, important knowledge about the health status of the plants and equipment can be lost. This could lead to unexpected stops during the production year. Manual analysis of condition monitored data may not yield maximum benefits as it might be difficult to notice failure patterns due to small changes that cannot be noticed through paper based inspections.

The results of this paper indicate that;

- eMaintenance is not fully utilized in this mining company.
- Apart from vibration analysis which has been outsourced, the mining company does not have an eMaintenance system that can analyse the other data it gathers through inspections.
- There is the lack of knowledge to analyse the data gathered through condition monitoring hence the

physical condition and the RUL of the are not always known.

The performance of the plant would be extensively improved if these issues are adequately considered in a revised Maintenance program. The outcome of this could be reduced unplanned stops as well as its economic, safety and operational consequences. Future studies could look into making hand held devices work with the current level of automation and computerized systems especially for smart phones such as the windows mobile handsets in an eMaintenance package.

6. REFERENCES

- [1] oubray, J. (1997). Reliability-Centered Maintenance. Industrial Press Inc.
- [2] Zio, E. (2009). "Reliability engineering: Old problems and new challenges". Reliability Engineering & System Safety, 94(2), pp. 125-141.
- [3] Muller, A., Crespo M. A., and Jung, B. (2008). "On the concept of e-maintenance: Review and current research". Reliability Engineering & System Safety, 93(8), pp. 1165-1187.
- [4] Mobley, R. K. (2002). An Introduction to Predictive Maintenance.
- [5] Campbell, J. D. (1995). "Outsourcing in maintenance management: A valid alternative to self-provision". Journal of Quality in Maintenance Engineering, 1(3), pp. 18-24.
- [6] Murthy, D., Atrens, A. and Eccleston, J. (2002). "Strategic Maintenance Management". Journal of Quality in Maintenance Engineering, 8(4), pp. 287-305. 2002.
- [7] Cross, M. (1988). "Raising the value of maintenance in the corporate environment". Management Research News, 11(3), pp. 8-11.
- [8] Altmannshoffer, R. (2006). "Industrielles FM. Der Facility Manager (in German)", April Issue pp. 12-13.
- [9] Alsayouf, I. (2007). "The role of maintenance in improving companies' productivity and profitability". International Journal of Production Economics, 105(1), pp. 70-78.
- [10] Kobbacy, K. A. H. and Murthy, D. N. P. (2008). Complex System Maintenance Handbook. Springer London.
- [11] O'Day, D. K. and Neumann, L. A. (1984). Assessing Infrastructure Needs: The State of the Art.
- [12] Utne, I., Brurok, T. and Rodseth, H. (2012). "A structured approach to improved condition monitoring". Journal of Loss Prevention in the Process Industries, 25(3), pp. 478-488.
- [13] Raheja, D., Llinas, J., Nagi, R. and Romanowski, C. (2006). "Data fusion/data mining-based architecture for condition-based maintenance". International Journal of Production Research, 44(14), pp. 2869-2887.
- [14] Si, X., Wang, W., Hu, C. and Zhou, D. (2011). "Remaining useful life estimation – A review on the statistical data driven approaches". European Journal of Operational Research, 213(1), pp. 1-14.
- [15] Kajko-Mattsson, M., Karim, R. and Mirjamsdotter, A. (2010). "Fundamentals of the eMaintenance concept". Presented at 1st International Workshop and Congress on eMaintenance.

- [16] Han, Y. and Song, Y. H. (2003). "Condition monitoring techniques for electrical equipment-a literature survey". *Power Delivery, IEEE Transactions*, 18(1), pp. 4-13.
- [17] Mitchell, J. and Carlson, J. (2001). "Equipment asset management—what are the real requirements?". *Reliability Magazine* 4 pp. 14.
- [18] Schuman, C. A. and Brent, A. C. (2005). "Asset life cycle management: Towards improving physical asset performance in the process industry". *International Journal of Operations & Production Management* , 25(6), pp. 566-579.
- [19] Dunn, S. (2009). "Condition monitoring in the 21st century," The Plant Maintenance Resource Center, Assetivity Pty Ltd,
- [20] Arts, R. Knapp, G. M. and Mann, Jr., L. (1998). "Some aspects of measuring maintenance performance in the process industry". *Journal of Quality in Maintenance Engineering*, 4(1), pp. 6-11.
- [21] Campos., J. (2009). "Development in the application of ICT in condition monitoring and maintenance". *Computers & Industrial Engineering*, 60(1), pp. 1-20.
- [22] Al-Jumaili, M., Rauhala, V., Tretten, P. and Karim, R. (2011). "Data quality in eMaintenance: A call for research". Presented at International Conference on Maintenance Performance Measurement & Management: MPM².
- [23] Candell, O., Karim, R. and Söderholm, P. (2009). "eMaintenance—Information logistics for maintenance support". *Robotics and Computer-Integrated Manufacturing*, 25(6), pp. 937-944.
- [24] Tsang, A. H. (2002). "Strategic dimensions of maintenance management". *Journal of Quality in Maintenance Engineering* 8(1), pp. 7-39.
- [25] Prakash, O. (2006). "Asset management through condition monitoring - how it may go wrong: A case study," In *Proceedings of the 1st World Congress on Engineering Asset Management*.
- [26] Fernandez, O., Labib, A. W., Walmsley, R. and Petty, D. J. (2003). "A decision support maintenance management system: Development and implementation". *International Journal of Quality & Reliability Management*, 20(8), pp. 965-979.
- [27] Biswal, M. and Parida, A. (2010). "An integrated approach for open e-maintenance: Opportunities and challenges". *Proceedings of eMaintenance*, pp. 22-24.

eMaintenance 3A

Computerised Analysis of Text Entry Fields in Maintenance Work Orders Data

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

Mustafa Aljumaili

Division of Operation, Maintenance
and Acoustics

Luleå University of Technology
Luleå, SE-971 87, Sweden
+46 920 49 15 53

mustafa.aljumaili@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

Enterprise resource planning systems and computerised maintenance management systems are commonly used by organisations for handling of maintenance work orders through a graphical user interface. A work order consists of a number of data fields, such as drop-down lists, list boxes, check boxes and text entry fields. In contrast to the other data fields, the operator has the freedom to type in any text in the text entry fields, to complement and make the work order description complete. Accordingly, the text entry fields of work orders can contain any words, in any number, as necessary. Data quality is crucial in statistical analysis of work orders data, and therefore manual analysis of work orders' text entry fields is often necessary before any decision making. However, this may be a very tedious and resource consuming process. In this article, we apply computerised analysis of text entry fields of work orders data, to study if it can bring further value in the assessment of technical assets' performance.

Keywords

Data quality, eMaintenance, maintenance, work orders, failure, decision support

1. INTRODUCTION

Maintenance can be described as the combination of all technical and administrative actions, including supervision actions, intended to retain an item in, or restore it to, a state in which it can perform a required function [2, 3]. Maintenance can be divided into preventive and corrective maintenance, where both generally are followed up regarding performance and costs. Information technology (IT), such as enterprise resource planning (ERP) systems and computerised maintenance management systems (CMMS), are used for such activities. The data of preventive and corrective maintenance work is commonly called work order data, following a set template and procedure for registration and closure, through a graphical user interface (GUI). Work orders contain a number of fields/boxes, such as: work order identification number; asset information regarding system, subsystem and components; maintenance activity; failure cause; and remedy. However, it depends if it is corrective or preventive maintenance work orders. The work order fields within a GUI comprise of drop-down lists, list boxes, check boxes and text entry fields. In contrast to the other data fields, the text entry fields are filled as the operator thinks is necessary for the

understanding of the work carried out. Accordingly, the text entry fields of work orders can contain any words, in any number.

High quality information is dependent on the quality of the raw data and the way in which it is processed. This fields' information is objective and depends on the user opinion. Since data processing has shifted from the role of providing operations support to being a major aspect of the operations themselves, [8]; therefore, analysis of this data is important for decision support.

For monitoring of maintenance performance and costs, computerised analysis and eMaintenance solutions [5, 6] are applied by organisations, and especially in asset intensive or safety oriented organisations, e.g. manufacturing, transportation, aviation and nuclear. However, manual data analysis of work orders' text entry fields is normally required before any decision making, which can be a tedious and resource consuming process.

In addition, data quality issues are mostly related to manual input and human errors. About 80% of data quality issues are related to human errors, while only 20% are related to machine failures [1]. Therefore, this process can be simplified by automated analysis of the text entry fields, and thus, the data quality can be increased, requiring less manual intervention.

In this article, we apply computerised analysis of work orders' text entry fields, to study if it can bring further value in the assessment of technical assets' performance, by relating text entry field data to other data fields. After describing the methodology, a case study is carried out on linear assets, and more specifically on railways. However, the methodology is generic and similar for other technical assets and organisations

2. METHODOLOGY

Lack of relevant data and information is one of the main problems for decision-making within the maintenance process [6]. The provision of the right information to the right user with the right quality and at the right time is essential [4, 6]. High-quality data are commonly defined as data that are appropriate for use [7, 8]. Wang [8] presents a framework of data quality consisting of four categories: intrinsic, contextual, representational and accessibility, as described in Table 1.

Furthermore, Wang [8] describe representational data quality as aspects related to the format of the data (concise and consistent representation) and meaning of data (interpretability and ease of understanding). Consequently, an input field which is free to fill out, i.e. text entry fields, results in data quality issues, at the same

time as work orders' text entry fields are vital as predefined fields cannot describe all possible events.

Table 1. Data quality aspects [8]

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

For analysis of text entry fields in work order data, an algorithm was written using MATLAB. The main steps of the algorithm, which is basic in programming, are shown in Figure 1.

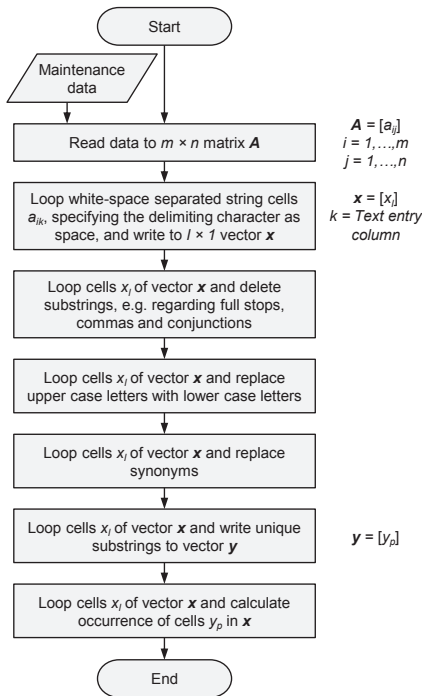


Figure 1. Flowchart.

The number of occurrence of these unique words can provide information about failure causes, types of failures, and the items that have more failures than others. Therefore, these extracted words, found to be interesting, can then be compared and linked to analysis of the other data fields, for additional information and study of agreement, or disagreement.

3. CASE STUDY

A case study has been carried out on railways to validate the methodology discussed. The data used in this study was provided by Trafikverket (Swedish Transport Administration). However, analysis in other organisations and for other assets will not yield the same result, but the methodology for assessing can be the same.

3.1 Data Collection

Operation and maintenance data has been collected from the Swedish railway section 111. Section 111 is a 128 km 30 tonne axle load mixed traffic section of the Swedish Iron Ore Line, stretching from the border of Norway, Riksgränsen, to Kiruna city (Figure 2).

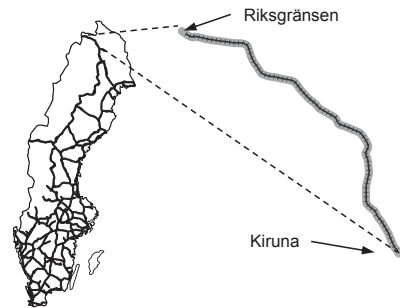


Figure 2. Swedish railway section 111 stretching from the border of Norway, Riksgränsen, to Kiruna city.

The failure data is collected from Trafikverket and constitute of infrastructure related corrective maintenance work, i.e. failure data. The corrective maintenance 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 failure data is from 2001.01.01 – 2014.01.01, i.e. 13 years, which in total gives 10 958 work orders. The main types of train delaying corrective maintenance failures are shown in Figure 3.

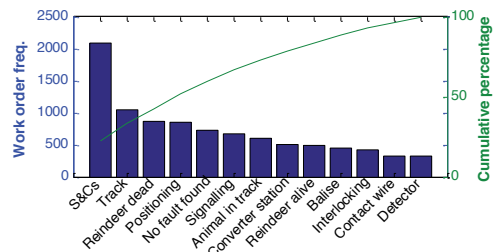


Figure 3. Types of failures of Section 111. S&Cs equals switches and crossings.

3.2 Work Order Data Quality

Simple data quality checks have been carried out on the failure data (Figure 4). Each work order consists of 71 fields. Fields with 100 % usage means that all work orders have some text or numbers filled in. Therefore, a data field with low usage can mean that the data quality is low, or it may just not be applicable for every work order. However, some data fields may be missing during the input process due to an error. Hence, this figure provides an idea about the data quality of the work orders, and how this missing data may affect the decision based on this data. The figure gives information regarding which data of the work orders are suitable for case studies. Also, in this way it is possible to improve the work order process, e.g. by removing unnecessary fields and improving the way of completing other fields.

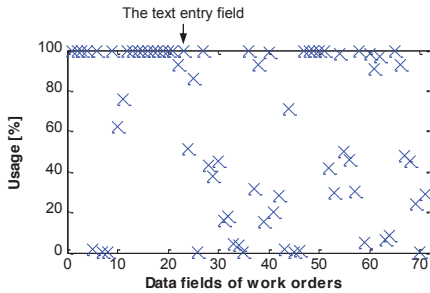


Figure 4. Usage of fields in the failure work orders.

The text entry field for describing the failure and work carried out is used in more than 99 % of the work orders (Figure 4). Thus, further analysis can be done as the field is frequently used and its information can vary depending on the user description.

3.3 Results and discussion

The comments field of the 10 958 work orders is found to contain 69 382 words in total. After removing commas, full stops and changing capital letters to lower case, the number of unique (disregarding repetitions) words is found to be 8 442; see Table 2.

Table 2. Usage of the text entry field.

Total number of words in the text entry field, i.e. the description field	69 382
Number of unique words, i.e. disregarding repetitions	11 400
Number of unique words after removing commas and full stops	9 756
Number of unique words after changing capital letters to lower case	8 442

By sorting the unique words by occurrence, we noticed that the most used words are found in more than one thousand work orders. Furthermore, by limiting the study to the 250 most used words, it is seen that the 250th word occurred in 39 work orders, i.e. not many in comparison to the total number of work orders. After removing needless words, e.g. conjunctions and prepositions, 143 unique words are left of the 250. Finally, through grouping of similar words, e.g. singular, plural and

synonyms, we end up with 104 words. Figure 5 shows the first 80 unique words with the highest occurrence. The terms are translated from Swedish to English, and consequently, the translated terms can consist of several words, like “error code” is one word in Swedish. A number of terms are marked out by arrows in Figure 5 for discussion.

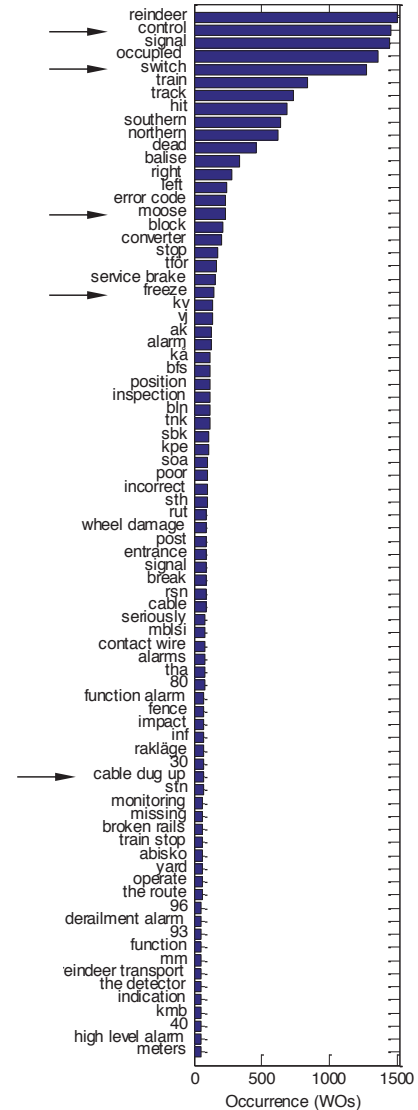


Figure 5. Occurrence of unique words in the text entry field of maintenance work orders data.

The second highest occurring term, "control", means that switch points are not in control/position, and thus, the term "control" could be aggregated with the fifth most used term, "switch", which then would become the most occurring term. By comparing Figure 5 with Figure 3, it can be seen that the top failure types and terms' occurrences are similar.

The term "moose" is found on the 16th place, occurring in 234 work orders (cows, bulls and calves included). By studying these 234 work orders, it is found that it would be hard to manually identify these work orders from the 10 958 work orders. The data can manually be sorted on animals in track and on the animal moose, but it would only give 149.

Another term is "freeze", which occurred 144 times. Freeze is referring to computer freeze/hang. By studying the 144 work orders, it is found that it would not be possible to sort them out without reading the free entry field of each of the 10 958 work orders, i.e. 69 382 words.

The next term is "cable dug up" (often costly), occurring in 68 work orders. By manually sorting the 10 958 work orders for cable systems, 172 work orders would be found, which would include 47 of the 68 cable dug up. In other words, computerised analysis gives additional information.

Lastly, the terms "96" and "93" (not marked with arrows) are identification of trains, which makes it possible to compare which trains are linked to most railway infrastructure failures.

Results from the computerised analysis give statistics on the occurrence of unique words in work orders' text entry field, i.e. the descriptive text typed in by operators. Subsequently, the unique words have been used to extract work orders according to a specific unique word in the text entry field. The extracted work orders have then been compared by trying to extract the same work orders from the whole dataset manually, by use of filtering functions in spreadsheet software. However, spreadsheets are filtered by use of drop-down lists, or pivot tables, i.e. it cannot filter text entry fields. The comparison showed that some information can only be found in the text entry field, and consequently, manual reading of the text entry field is required to capture all work orders related to a specific type of failure or unique word. For the data used in this study, in a worst case scenario, it would mean reading text entry fields of 10 958 work orders, which equals 69 382 words. Alternatively, computerised analysis can be carried out, which is fast and simple when algorithms are in place. It is clear from this study that computerised solutions can save the time of data analysis. In addition, more information can be driven that may help to support decision making process.

4. CONCLUSIONS

A methodology for computerised analysis of text entry fields of maintenance work orders data has been described and studied in a case study. It has been found that such analysis makes the process more efficient, gives additional information, and in some cases, it is the only realistic methodology for analysis of maintenance

work order data, as long as resources for manual analysis is not endless. Moreover, since the described method improves identification of failures, it improves the input data to reliability and availability studies, as missing observations (failures) can have a large effect on the mean time between failures (MTBF) and maintenance times.

This analysis also can provide an overview about work order data quality as well, what data is missing, which data is more important regarding data usage. This may be useful for decision makers before making any decision based on this data.

Finally, it is necessary to mention here that, missing data may not mean that the data quality is poor. Further study about the data fields constraints and metadata should be conducted. In some cases, the input to these fields may be optional. Therefore, the final perception about data quality should be considered after checking all work order fields constraints and metadata.

5. ACKNOWLEDGMENTS

The authors would like to thank Luleå Railway Research Centre (JVTC) and Trafikverket (Swedish Transport Administration) for their support and funding of the research.

6. REFERENCES

- [1] Aljumaili, M., Tretten, P., Karim, R., Kumar, U.D., "Study of aspects of data quality in e-maintenance", *Int J Cond Monit Diagn Eng Manag*, 15(4), pp. 3-14, 2012.
- [2] CEN, "EN 13306: Maintenance Terminology", European Committee for Standardization, Brussels, 2010.
- [3] IEC, "IEC 60050-191: International Electrotechnical Vocabulary: Chapter 191: Dependability and quality of service", International Electrotechnical Commission (IEC), Geneva, 1990.
- [4] 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.
- [5] Muller, A., Crespo-Marquez, A., Jung, B., "On the concept of e-maintenance: review and current research", *Reliability Engineering and System Safety*, 93(8), pp. 1165-1187, 2008.
- [6] Parida, A., "Maintenance performance measurement system: Application of ICT and e-Maintenance concepts.", *International Journal of COMADEM*, 9(4), pp. 30-34, 2006.
- [7] Strong, D.M., Lee, Y.W., Wang, R.Y., "Data quality in context", *Communications of the ACM*, 40(5), pp. 103-110, 1997.
- [8] Wang, R.Y., Strong, D.M., "Beyond accuracy: What data quality means to data consumers", *Journal of Management Information Systems*, 12(4), pp. 5-34, 1996.

Control Charts supporting Condition-Based Maintenance of Linear Railway Infrastructure Assets

Bjarne Bergquist
Luleå University of Technology
Universitetsvägen 1
SE-971 87 Luleå, Sweden
+46 920 49 21 37
bjarne@ltu.se

Peter Söderholm
Trafikverket (Swedish Transport Administration)
Box 809
SE-971 25 Luleå, Sweden
+46 76 792 53 85
peter.soderholm@trafikverket.se

ABSTRACT

This paper presents a control chart approach for monitoring, diagnostics and prognostics to support condition-based maintenance (CBM) using condition data of linear railway infrastructure assets. The condition data were obtained from regular inspections done by a railway track measurement wagon. The condition data were statistically analysed by using two different control charts to evaluate the possibility for earlier detection of derailment hazardous faults using both temporal and spatial information. The study indicates that the proposed control chart approach can be used for condition assessment of track and thereby provide valuable decision support for CBM. The control chart for condition information in the temporal domain supports diagnostics, while the control chart for condition information in the spatiotemporal domain also supports prognostics. The two proposed control charts give earlier fault warnings compared to the traditional approach. This facilitates decisions regarding CBM actions with an extended planning horizon and gives the possibility to increase the operational availability of track.

Keywords

Condition assessment, control chart, spatiotemporal, condition monitoring, diagnostics, prognostics, condition-based maintenance, decision support, maintenance limits, railway infrastructure, linear assets, railway track, Sweden.

1. INTRODUCTION

Today, there are many maintenance approaches where condition-based maintenance (CBM) is central, e.g. predictive maintenance [1], maintenance excellence [2], prognostics and health management (PHM) [3], integrated vehicle health management (IVHM) [4], and e-Maintenance [8-12]. One prerequisite of CBM is the concepts of functional failure (a faulty state) and potential failure (a failure event) [5]. This also necessitates the description of system functions in relation to stakeholder requirements [7]. A required function is a function or a combination of functions of an item, which is considered necessary to provide a given service [6]. Hence, CBM relies on continuous monitoring and intermittent test or inspection of the condition of an item, which then is assessed to achieve diagnoses of the item's current condition and prognoses of its future condition [7].

As argued by Nowlan & Heap [5] and their successors, maintenance limits based on the rate of degradation (potential

failure or failure event), rather than the breach of a critical limit (functional failure or a faulty state) would reduce the level of emergency, producing earlier alarms and increasing possibilities of planned preventive rather than acute corrective maintenance. However, selecting these earlier maintenance limits in a systematic way, while balancing the risk of undetected safety-critical faults and false alarms is challenging [1, 5, 7]. In addition, the nature of collected condition data usually violate common statistical assumptions, such as samples being taken from an independently, identically distributed (IID) normal distribution. One example is condition data or quality characteristics sampled along, say, a railway track. For example, if the track gauge (spacing of the rails on a railway track), when measured (between the inner faces of the load-bearing rails), is found to be too large at a particular location, the track gauge is likely about as large if a new measurement is taken one meter, or even ten meters from the first measurement. When measurements are taken near each other, the measurement data are therefore dependent, or autocorrelated, and therefore violate the independence assumption. Furthermore, many measurement data also deviate from the Gaussian distribution, i.e. they are not normally distributed.

Naturally, the violations of the common statistical assumptions given above differ from being of no practical consequence to invoking serious errors. Hence, it is necessary to investigate the severity of violations so that proper analytical approaches that master, for instance data sampled from non-symmetric distributions, and dependencies in the data may be applied.

The purpose of this paper is to describe a proposed control chart approach for condition assessment of linear assets. Linear assets are those that from many points of view could be described as two-dimensional, e.g. track and catenary systems within railway. The location of a point or a section along the asset could for instance be defined by its distance from the asset's known endpoint. A maintenance work order, e.g. tamping or welding of the track, would typically include a location or an interval given by its specific distance from a station or another given point. Based on the stated purpose, the following research question was formulated: How can statistical approaches, statistical process control in particular, be applied to assess the condition of linear assets? The approach is illustrated by a case study of track twist failure at the Swedish Iron ore line.

2. STATISTICAL CONDITION MONITORING USING CONTROL CHARTS

Condition monitoring means collecting data that represent the system's condition in some way [1; 2; 13]. Diagnostics is concerned with the interpretation of collected condition data and the conclusion drawn about the system's current condition [13]. These conclusions are then used to make decisions regarding CBM [1; 2; 14; 15]. An extension of diagnostics is prognostics, which tries to predict the future condition of a technical system [1; 16; 17]. The aim of prognostics is to stop critical functional failures before they occur [1; 16; 18]. The prognostic information also enables decisions about recommended albeit non-acute maintenance that is advantageous to perform along with currently required maintenance [1; 14; 15]. In addition, choices about continued operation, with or without any restrictions, or terminated operation could be based on the diagnostic information [5]. Similarly, decisions about the operation can also be based on prognostic information, with the advantage of a planning horizon [1; 15]. Hence, prognostics also enable a control of the aging of technical systems, which may be required by regulatory authorities [17].

One challenge with prognostics is that the location of new observations coming from fully stochastic processes cannot be precisely predicted. However, many types of deterioration behaviours are largely deterministic. Hence, prognostic models can be used to predict item failures, given that the item conditions and loads can be measured or estimated. Statistically-based prognostic models are regularly used for making prognoses related to items where deterministic and stochastic behaviours coexist. However, the quality of the prognoses depends on many aspects such as the quality of the condition data fed into the models, the quality of the diagnostic and prognostic models themselves, and the degree to which the mechanism to be predicted is deterministic, chaotic, or stochastic.

Statistical process control (SPC) is a classical statistical approach used for many surveillance applications to monitor processes. The SPC approach was originally developed in the 1920s [19], but has since then found use in various sectors [20]. SPC is based on control charts, where measurements of the process are monitored and compared to control limits based on the statistical distribution from which the data is assumed to be sampled from. An observation is classified as being within its expected range if it remains within the control limits. However, if the observed data point is outside of the control limits, it is reasonable to assume that the process is affected by systematic variation and needs attention.

Many types of control charts have been designed for various purposes. For some processes, e.g. within manufacturing, it is convenient to sample and measure variables in groups, so called rational subgroups. Automatic measurements, for instance measuring all products or continuous monitoring of some quality characteristics are, however, increasingly common. Other control charts include those suitable for various situations, such as when the data is categorical or numeric, for individual or multivariate properties, for skewed distributed data and so on [21]. There are also examples of control charts that have been used to establish predictive maintenance plans. One example of the latter is Katter et al. [22], who use control charts to monitor laser equipment to establish CBM of the cathode. Another maintenance-related

application is described by Ben Daya and Rahim [23], who suggest that control charts could be used to monitor processes where in-control periods are followed by periods with increasing failure rates. The maintenance-related data used in this paper are variables obtained using automatic sampling, which is suitable for control charts for individual observations. Hence, the discussion is from this point restricted to control charts for such data.

The selection of control limits is based on balancing the risks of not detecting an assignable cause, the beta risk, with the alpha risk: erroneously indicating an assignable cause. Setting control limits that are too wide will increase the beta risk, i.e. result in undetected failures. On the other hand, too narrow limits will increase the number of false alarms, i.e. the alpha risk. Furthermore, these testability deficiencies at single test levels in combination with insufficient integration between different test levels (e.g. test during operational vs. maintenance settings) may result in no fault found (NFF) and dead-on-arrival (DoA) events [7, 24]. 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 generate false alarms in 1 out of 370 observations.

The detection capability can be described in a similar manner. If an assignable cause was to shift the mean value of the process by 1σ , a regular so called individuals x-chart with control limits of 3σ would have a 2.3% chance of detection already at the first observation of the process after the shift, and a 50% chance of detection of the assignable cause generating a deviation as large as 3σ from the nominal value. See also Montgomery [21].

When dealing with linear assets, the challenge of achieving good testability increases. The track is stressed by axle loads ≥ 5 metric tonnes from the measurement wagon itself, while the entrepreneur trying to localize and correct faults relies on unstressed measurements, which will differ from the stressed counterparts. These different test levels may lead to NFF. Similarly, positioning errors between consecutive measurements of the same part of a liner asset may result in NFF events. This deficiency may be seen as insufficient testability due to incorrect spatial fault localisation in the time domain. [7, 24]

3. STUDY APPROACH AND CASE STUDY

Based on systematic selection criteria [25] (i.e. type of research question, no required control over behavioural events, and focus on contemporary events, but also criticality and extremeness of the case) a single case study of the Swedish Iron ore line was chosen as an appropriate research strategy to answer the stated research question.

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 and control charts. Finally, the paper has been reviewed by key informants and roles to verify its content.

The Iron ore line is a bottleneck in the mining companies' logistic chain. Hence, the availability of the line is essential. To minimize transport disruptions, maintenance of vital items of the railway infrastructure should therefore be preventive and condition-based instead of corrective. The most critical linear assets of the railway infrastructure are the catenary system and the track. This study focuses on track, since its condition is fundamental to the railway

system, where track failures potentially affect safety and may cause delays due to speed restrictions or derailments. More specifically, twist failure of track was selected based on its criticality with regard to safety.

The track condition data used are collected by a measurement wagon. The wagon is regularly pulled along the track system in speeds up to 200 km/h and measures each section of the Swedish track system up to six times per year depending on the section's criticality. Observations of about 30 track geometry variables are obtained and stored for every 25th cm. Track variables include the position coordinates (height, and locations in the plane), and the track width [26; 27].

One critical track geometry variable is the cant, which 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. On a 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. However, there is a risk of derailment if the cant changes too rapidly. This phenomenon is called twist, i.e. the rate of change of the track superelevation. The twist is defined as the algebraic difference between two cants taken at a defined distance apart, usually expressed as a gradient between the two points of measurement, i.e. expressed as a ratio (% or mm/m). Twist measurements is either taken simultaneously at a fixed distance, e.g. at a distance equivalent to the wheel-base, or is computed from consecutive measurements of cant. Normally, the twist is measured on a 6 m base, i.e. the cant measured at two points with 6 m distance.

The measurement data is stored in a database (Opram [28; 29]) together with information about when and where the measurements were performed. The Opram database also contains information about the infrastructure and its attributes (e.g. type of object, geographical position, and description) and if the measurement is taken on a point asset (e.g. railway switch and level crossing) or a linear asset (e.g. track and catenary system). The database also contains information about events and their history, e.g. track alignment and related information. The Opram 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.

4. PROPOSED CONTROL CHART APPROACH FOR DIAGNOSTICS AND PROGNOSTICS

From a maintenance perspective, the differentiation between point assets and linear assets is depending on the criticality that the length of the asset has. The length of a point asset is not critical for its maintenance, e.g. a railway switch, a level crossing or a way-side detector for monitoring of the rolling stock. When dealing with a point asset, maintenance actions are not assigned to a particular length of the asset, but rather to the entire asset or to some of its indenture levels (included items). However, a linear asset is an asset whose length plays a central role in its maintenance, e.g. railway track and catenary system. When maintaining linear assets (grinding, tamping, welding et cetera), it is necessary to be able to define the location of a point or a section along the asset.

From a statistical point of view, 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 at the same part of the railway infrastructure may also exhibit autocorrelation in the time domain. For example, the measurement of 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 been subjected to normal usage and not to repair or abnormal and mainly stochastic events such as accidents or extreme weather.

Generally, autocorrelation of control charts data may be handled through simply removing nearby observations until the autocorrelation is low enough to not cause concerns. When the ACF plot shows that the autocorrelation is insignificant at, say, lag 5, removing four out of five consecutive observations will generate a data series that can be analysed using regular control charts. This route could not be used for the spatial data since the autocorrelation is strong several hundred observations apart. A removal of the data necessary to remove autocorrelation would also make the chart too blunt for the purpose of locating failures along the track. Autocorrelation can also be handled using two other distinct routes when applying control charts [30]. One route is to plot the residuals of a time series model on a standard control chart, and the other route is to adjust the control limits to compensate for autocorrelation. In this paper, the latter route is used. The control limits were based on standard deviations of the spatial data from a time series model of a large sample so that the collected data include all naturally occurring variation of the measured property. In Figure 1, a time series plot of the 6m twist is shown of a track section including more than 65 000 observations, which equals data from a 16.5 km long section.

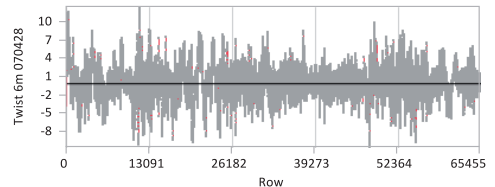


Figure 1. Time series plot of 6m twist obtained April 28, 2007

As seen in Figure 1, the variation is not constant, however, it is still assumed that the process is in statistical control. Hence, it is assumed that there are no known and assignable causes for variation, and thus the data can be used for calculation of the distribution properties of the process. The standard deviation calculation is particularly interesting, since small sample calculations of highly dependent data, such as these, would underestimate the total variation in the data. Using time series analysis, the estimated standard deviation was 2.42 mm. As mentioned earlier, curves are designed with a controlled change of cant due to operational requirements of the rolling stock, which becomes part of the natural twist distribution. However, larger twists are normally due to some geometrical deficiency in relation to the intended infrastructure design. The distribution deviates

significantly from a normal distribution, with higher density both in the tails and around zero, see Figure 2. Deviations with increased probability density in tails have been shown to reduce the performance of individuals' charts [31].

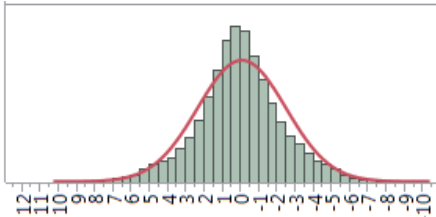


Figure 2. Distribution of twist data [mm] from April 28, 2007. Continuous curve represent fitted normal distribution

Control charts are generally used to test for out-of-control conditions of processes, but it is necessary to select appropriate control limits before a control chart is created. A common choice is to use three standard deviations, which, given that the data are normally distributed and that the distribution properties (mean and standard deviation) were known would generate a risk of false alarm around 1/370. Empirically-based control limits were used in this study since the distribution was found to be non-normal, and the 0.135% and 99.865 % percentiles together contain the probability proportion of 1/370. In this case, the percentages of these tails differ slightly (-8.23 and 8.26) due to the low frequency of observations in the tails. Hence, the average of these empirical percentiles was considered to be a better representation of the distribution, and therefore the control limits were set to ± 8.25 .

5. RESULTS

In this section, two different proposed control chart approaches for assessment of the track twist condition for diagnostic and prognostic purposes that supports decision making related to CBM are outlined.

5.1 Diagnostics supported by a temporal control chart approach

A derailment hazardous twist was detected on a section of the Iron ore line in June 10, 2011 by using the traditional alarm limits. The hazardous twist was found at track section 111, between marker 1495 and 1496, which is west of Kiruna. This alarm was used as a starting point for a further statistical analysis of the same section using both that measurement data, as well as data from both earlier and later measurements. This new analysis was performed by using ordinary Shewhart type individuals control charts, based on the empirical percentile control limits.

It is clear from the charts in Figures 3 and 4 that irregular twisting would have been detected earlier using a control chart approach than by the traditional practice relying on safety-related alarm limits based on geometrical properties (in this case 25mm). The charts would signal for an assignable cause at least three months earlier (April 2011, Figure 4) than the measurement requiring immediate actions to adjust the track positions. The difference in twist locations are due to erroneous positioning data in one or both the two measurements. Range charts are not shown due to the large spatial autocorrelation.

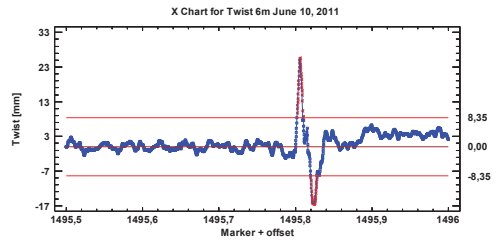


Figure 3. Twist on railway at marker 1495, June 2011

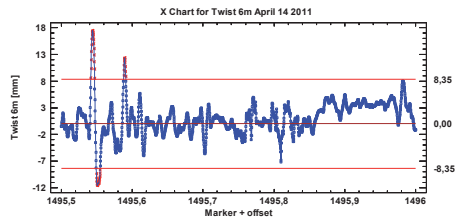


Figure 4. Twist on railway track section 111, near marker 1495, April 14, 2011

5.2 Prognostics supported by a spatiotemporal control chart approach

Repeated measurements make temporal studies of the track section possible [32]. However, a temporal graph requires that the position markers of each measurement are comparable, and as seen from Figure 3 and 4, the positioning error was large in relation to the wavelengths of the track twist (in this case 300 m). The studied track section was therefore split into 300 m intervals to overcome the positioning error and enable monitoring of the change in twist by using successive passages of the measurement wagon. The twist error can be considered similar to a short wavelet function appearing along the track; a negative twist for instance due to that one rail has sunk, must be followed by a positive twist when the sunken rail rises back after the deformed section has been passed, and even rises past the previous base level, due to the stiffness of the rail. A slight positioning error between two consecutive measurements could therefore generate a strong positive twist at a certain position, followed by zero or negative twist at the seemingly same position at the next measurement occasion.

A data binning procedure was utilized to overcome the positioning error. 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 is therefore used here as a measure of twist problems. The range of the twist variation was measured within each 300 m section, assuming that the range of twist within such a section would be a good measure of track twist problems in the section.

Box-Cox transformation test of the twist ranges suggested that the range values should be transformed, and suggested 95% confidence interval for the power constant between -0.27 and -0.00. 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 the range is skewed to the right and the minimum range has zero as a natural lower limit.

The standard deviation and mean of the logarithmic twist was estimated using data from 17 in-control periods from track section 111, marker 416 to 417.

In Figure 5, four 300 m sections tracking another twist error 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. Here, the 'product' represents a track section and the repeated measurements are observations from the consecutive measurements; the oldest (April 28, 2007 near the left section border and the latest, September 21, 2012 to the right in each section).

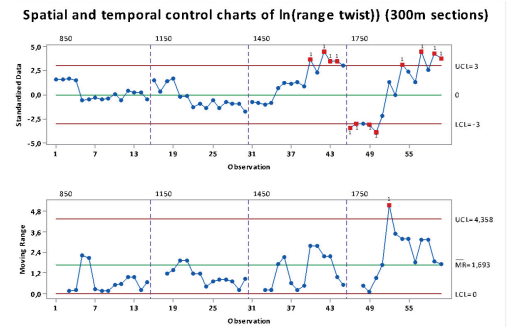


Figure 5. Logarithm of range of twist of track divided into 300 m sections. Each observation represents one measurement; the oldest to the left of one section

Each 300 m section is delimited by the vertical dashed lines. Spatial and temporal information from 18 consecutive measurements taken between April 2007 and September 2012. The observations in Figure 5 have then been scaled to unit variance and centred to zero. The Z-chart used in this way thus lets the analyst study both spatial and temporal data simultaneously. The top chart of Figure 5 shows how a certain measurement obtained by the measurement wagon deviates, and it is seen that several observations generate alarms. In this case, the first alarm would be had 26 August, 2011 on the 1450 section, and June 11 the same year on section 1750. The regular procedure with set alarm limits indicated a derailment hazardous twist fault on October 6, 2011. Note also that the moving range chart (here based on three consecutive measurements) signals for section 1750 even earlier, the 16th October 2010. This means that the chart has reacted on the rapid change of the twist ranges at this section. In this study, no control charts for spatiotemporal data where the spatial information is one-dimensional as the track data is has been found. However, spatiotemporal data using two-dimensional information has been suggested for outbreak detection [33] and for monitoring using image analysis [34]. Depending on assumptions, the variation of the different products of a Z-chart can be estimated separately for each product, pooled, or scaled.

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. The figure also shows that the track demonstrates twist problems in the neighbouring section, see Figure 5. A final note about the chart: It is apparent that whatever actions that were taken to correct the twist, they did not fully restore the track, as the three consecutive measurements taken after October 6 also had twist problems. The observations in the moving range chart are usually well below the alarm limits, due to the temporal autocorrelation present.

6. DISCUSSION AND CONCLUSIONS

Two different control charts for plotting condition data are presented, i.e. in the spatial (Shewhart type control chart) and in the spatiotemporal (Z-chart) domains. The study illustrates how the use of the control charts in respective domain can improve maintenance performance compared to the current practice. This is achieved by the use of statistically-based alarm limits instead of relying on safety-related limits that are based on geometrical specifications. The spatial approach mainly supports enhanced diagnostics, while the spatiotemporal approach also supports prognostics. In addition, the proposed control chart approach also supports improved testability by providing a systematic approach to manage the risk of undetected failures and false alarms by balancing the alpha and beta risks when designing the control chart alarm limits. The procedure also has the potential to reduce NFF events through combining consecutive measurements and thus reduce the risks of single measurements having large positioning errors. Such large positioning errors would be easily spotted in the moving range charts of the spatiotemporal control chart (Figure 5).

The easiest implemented control charts of the two proposed is the one for diagnostic purposes, and it is also the one recommended for the studied case. The reasons for these recommendations are mainly two. The first reason is related to the positioning error of the measurement wagon, making analyses of deterioration dynamics based on repeated measurements challenging. The second reason for recommending the diagnostic chart is its relative simplicity, to limit the time needed for practitioners to work with charts of their own. However, given that consecutive measurements could be more easily obtained, the spatiotemporal approach holds greater potential, since it offers a visualization of the dynamic events of the asset degradation, which would support prognostics in addition to diagnostics.

However, to get acceptance for statistical process control among practitioners, a seemingly endless list of alarms may not be the best first outcome of the procedure. A control chart of a longer section of the track does reveal multiple positions where the twist would reach beyond the statistical limits, and maintenance budgets are always limited. Thus, statistical significance should be complemented with practical significance. Physical assets will naturally degrade, but maintenance actions should focus on those faults that are most severe, or rapidly becoming grave. For surveillance purposes, the alarm limits should therefore be set based on static statistics, on what is practical, and preferably also based on system dynamics, i.e. degradation.

Even though the approach used here may be considered conservative due to the inclusion of out-of-control points for estimating the natural variation, the control limits are narrower than what is practically feasible to act on. In cases such as this,

reasonable control limits must be set taking into account maintenance budgets and engineering, geological and hydrological know-how, in this case e.g. special soil conditions, areas of heavy rainfalls and so on. How to balance such know-how into a maintenance approach is, however, beyond the scope of this paper.

Even so, the proposed analysis procedure was successful in detecting the failure much earlier than the traditional 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 is an interesting continuation of this work.

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. One consideration is also practical consequences; the limits prompting actions may be found somewhere between the geometrically determined safety-related alarm limits and the more narrow statistically-based maintenance-related ones.

For this application, it is suggested that the current measurements of the track condition and geometric safety-related alarm limits should be complemented with statistical alarms. In this case, the geometric alarm limit for the 6m base twist is 25 mm, which corresponds to the inner wheel rim protrusion on the inside rail. Twists larger than 25mm are outing the train at derailment risk, since rims may lose ability to control the position of the wheels for cars without bogies. 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 maintenance action is needed. Statistical monitoring could complement the geometrically-based alarm limits for the latter purposes, given that statistical alarms have more narrow control limits compared to the geometrical methods. This complementary use of two sets of alarm limits supports a gradual implementation and fine-tuning of the proposed approach. The reason is that the more narrow statistically-based maintenance-limits not directly affects safety, but primarily availability performance and cost, which facilitates an implementation without a necessary involvement of responsible safety authorities.

Even though the proposed control chart approach is illustrated by twist failures of the track, it should be possible to 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 probably also be assessed by the proposed approach, e.g. the catenary system. However, besides in other contexts, further studies are needed to thoroughly study the performance of control charts for CBM. The possibility to use more advanced time series analysis for assessment of the condition of linear assets is also an area where this study has indicated interesting result and published research is lacking.

7. ACKNOWLEDGEMENTS

We gratefully acknowledge the intellectual and financial support given by Trafikverket (Swedish Transport Administration) and Luleå Railway Research Centre (JVTC).

8. REFERENCES

- [1] Mobley, R.K. (1990). *An Introduction to Predictive Maintenance*. Van Nostrand Reinhold, New York.
- [2] Campbell, J.D. & Jardine, A.K.S. (2001). *Maintenance Excellence: Optimizing Equipment Life-Cycle Decisions*. Marcel Dekker, New York.
- [3] Malley, M.E. (2001). *A Methodology for Simulating the Joint Strike Fighter's (JSF) Prognostics and Health Management System*. Master's Thesis, Department of Operational Sciences, Air University, Ohio.
- [4] Baroth, E., Power, W.T., Fox, J., Prosser, B., Pallix, J., Schweikard, K., & Zakrajsek, J. (2001). "IVHM (Integrated Vehicle Health Management) Techniques for Future Space Vehicles". In Proceedings of the 37th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit.
- [5] Nowlan, F.S. & Heap, H.F. (1978). "*Reliability-Centered Maintenance*". Report AD/A066-579, National Technical Information Service, US Department of Commerce, Springfield, Virginia.
- [6] IEV (2014). *International Electrotechnical Vocabulary (IEV)* Online Database. Available at: <http://std.iec.ch/iec60050>; Access Date: 24/04/2014.
- [7] Söderholm, P. (2005). *Maintenance and continuous improvement of complex systems: linking stakeholder requirements to the use of built-in test systems*. Doctoral thesis, Division of operation and maintenance engineering, Luleå University of Technology, Luleå.
- [8] Lee, J. (2007). E-manufacturing - fundamental, tools, and transformation. *Robotics and Computer-Integrated Manufacturing*; 19: 501-507.
- [9] Levrat, E., Iunga, B. Crespo Marquez, A. E-maintenance: review and conceptual framework. *Production Planning & Control* 2008; 19: 408-429.
- [10] Karim, R., Candell, O. Söderholm, P. Development of ICT-based maintenance support services. *Journal of Quality in Maintenance Engineering* 2009; 15: 127 – 150.
- [11] Karim, R., Kajko-Mattsson, M. Söderholm, P., Candell, O., Tyrberg, T., Öhlund, H., Johansson, J. Positioning embedded software maintenance within industrial maintenance. In: IEEE International Conference on Software Maintenance: ICSM 2008. IEEE, 2008; 440 – 443.
- [12] Karim, R., Kajko-Mattsson, M. Söderholm, P. 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. 75-80.
- [13] Martin, K.F. (1993). *Condition Monitoring and Fault Diagnosis with Applications to Machine Tool Subsystems*. Profitable Condition Monitoring. Kluwer Academic Publishers, Dordrecht, 99-111.
- [14] Litt, J., Simon, D.L., Meyer, C., DePold, H., Curtiss, J.R., Winston, H., Wang, Y., Statler, I., & Gawdiak, Y. (2000). NASA Aviation Safety Program Aircraft Engine Health Management Data Mining Tools Roadmap. Proceedings of Data Mining and Knowledge Discovery: Theory, Tools and

- Technology III, 24-25 April 2000, Orlando, Florida, 4057, 292-298.
- [15] Hess, A. & Fila, L. (2002). Prognostics, from the Need to Reality: From the Fleet Users and PHM System Designer/Developers Perspectives. Proceedings of the IEEE Aerospace Conference, 9-16 March 2002, Big Sky, Montana, 6, 2791-2797.
- [16] Becker, K.C., Byington, C.S., Forbes, N.A., & Nickerson, G.W. (1998). Predicting and Preventing Machine Failures. *The Industrial Physicist*, 4(4), 20-23.
- [17] Martorell, S., Sanchez, A., Munoz, A., Pitarch, J.L., Serradell, V., & Roldan, J. (1999). The Use of Maintenance Indicators to Evaluate the Effects of Maintenance Programs on NPP Performance and Safety. *Reliability Engineering & System Safety*, 65(2), 85-94.
- [18] Roemer, M.J., Nwadiogbu, E.O., & Bloor, G. (2001). "Development of Diagnostic and Prognostic Technologies for Aerospace Health Management Applications." In Proceedings of IEEE Aerospace Conference.
- [19] Shewhart, W. (1931). *Economic Control of Quality of Manufactured Product*, Van Nostrand Company, Inc. New York.
- [20] MacCarthy, B.L. Wasusri, T. (2002). A review of non-standard applications of statistical process control (SPC) charts. *International Journal of Quality and Reliability Management*; 19: 295-320.
- [21] Montgomery, D.C. (2008). *Introduction to statistical process control*. John Wiley & Sons, New York, 2008.
- [22] Katter Jr., J.G., Tu, J.f., Monacelli, L.E. Gartner, M. (1998). Predictive cathode maintenance of an industrial laser using statistical process control. *Journal of Laser Applications*, 10: 161-169.
- [23] Ben-Daya, M. Rahim, M.A. (2000). Effects of maintenance in the economic design of \bar{x} -control cart. *European Journal of Operational Research*; 1:131-143.
- [24] Söderholm, P. (2007). A system view of the No Faults Found phenomenon. *Reliability Engineering & Systems Safety*. 92(1): 1-14.
- [25] Yin, R.K. (2003). *Case Study Research: Design and Methods*, Sage, Thousand Oaks.
- [26] Banverket (1997). *BVF 587.02 - Spårlägeskontroll och kvalitetsnormer – Central mätvagn STRIX*. Banverket, Borlänge.
- [27] Banverket. (1997). *Z-647 97-11-15 – Beskrivning av spårlägesystemet i mätvagn STRIX*. Banverket, Borlänge.
- [28] Bentley (2012). *Bentley Optram*. Available at: <http://www.bentley.com/en-US/Products/Bentley+Optram/>, 20 August 2012.
- [29] Trafikverket. (2012). *Optram*. Available at: <http://www.trafikverket.se/foretag/bygga-och-underhalla/jarnvag/system-verktyg-och-tjanster-for-jarnvagsjobb/optram/>, 20 August 2012.
- [30] Stoumbos, Z.G., Reynolds, M.R., Jr. Ryan, T.P. Woodall, W.H. (2000). The state of statistical process control as we proceed into the 21st century. *Journal of the American Statistical Association*, 2000; 95: 992-998.
- [31] Vermaat, MB., Ion, RA., Does, RJMM., Klaasen, C.A.J.C. (2003). Comparison of Shewhart individuals control charts based on normal, non-parametric, and extreme-value theory. *Quality and Reliability Engineering International*; 19: 337-353.
- [32] Bergquist, B. & Söderholm, P. (2012). "Control Charts for Assessment of Linear Asset condition using both Temporal and Spatial Information". In Proceedings of: the 2nd International workshop and Congress on eMaintenance.
- [33] Jiang, W., Han, S.W., Tsui, W.-L. Woodall, W.H. (2010). Spatiotemporal surveillance methods in the presence of autocorrelation. *Statistics in Medicine*. 30: 569-583.
- [34] Megahed, F.M., Wells, L.J., Camelia, J.A Woodall, W.H. (2012). A spatiotemporal method for the monitoring of image data. *Quality and Reliability Engineering International*, 28: 967-980.

eMaintenance 3B

A conceptual database model for mobile e-maintenance

Jaime Campos
Linnaeus University,
Department of Informatics
SE-351 95,
Sweden
Jaime.campos@lnu.se

Erkki Jantunen
VTT Technical Research Centre of Finland
P.O.Box 1000
FI-02044 VTT, Finland
Erkki.jantunen@vtt.fi

ABSTRACT

A conceptual database model is the primary requirement for a physical database design. The paper discusses a conceptual database model, suitable for a manufacturing or process plant when a mobile device is utilised. A knowledge sharing and an e-learning module are suggested to be used together with the e-maintenance applications because of the benefits they can provide for the maintenance engineers when performing their different works tasks. Firstly, the authors briefly review the state of the art in the area. Then the e-maintenance approach and the information needs of the maintenance engineer are discussed as well as organisational learning and knowledge sharing are taken into account. Further, when the mobile database model is developed, it is recommended to apply the database standards of the MIMOSA (Machinery Information Management Open Systems Alliance) which is worthwhile considering for purposes of work order and measurement event. To conclude, there is a static view of a conceptual model illustrated, which highlights the entities and relationship for the surveillance of the machines as well as the most important e-learning parts, for this purpose, the web mobile technology is suggested for making use of geographically distributed databases through the Internet. These tools facilitate mobile e-monitoring and maintenance.

Keywords

E-maintenance, web technologies, condition monitoring, knowledge management, e-learning, database.

1. INTRODUCTION

The mobile devices within the mobile e-maintenance started to appear during 2006 where few applications were developed with the embedded technology [1]. In 2007 the first mobile device with web technology, i.e. ICTs, such as web and wireless technologies came forth, Campos et al. [2; 3], and the first with embedded technologies during 2006 [4]. There are other efforts on the mobile device applications for condition monitoring and maintenance as well as e-maintenance, such as (Jantunen et al. [5]; Irigaray et al. [6]; Campos et al [7]; Antonelli et al. [8]). The databases are an important part of any software application system and when preventive maintenance is applied, an immense amount of data on operations and maintenance is produced [9]. Data Base Management Systems (DBMS) have been applied to handle these data and records for different purposes in enterprises, including equipment life cycle management. The current paper proposes a database for

a mobile proactive maintenance system, i.e. an e-maintenance mobile system that supports the e-learning and knowledge sharing process between the employees in a manufacturing plant, i.e. the maintenance engineers and the expert. The combination of the e-maintenance and e-learning as well as knowledge sharing is important in maintenance, since the complexity of the work task of the maintenance engineer differs and depends on the fault occurred. It is, therefore, the use of the ICTs supporting the e-learning and knowledge sharing that becomes important as a part of the e-maintenance software solutions.

In the system development process of any software system these are important issues to consider before building, for instance, an e-maintenance decision support system, namely what kind of data to gather and how to conceptually model the data and manage their storage, how the data will be analysed and finally how an efficient data acquisition should be done. To be able to monitor a machine condition there is a need to gather a huge amount of data. The gathered data facilitate, for example, the determination of the deterioration trend of the machine. It provides the user with the possibility of analysing maintenance history and trends. Databases are important tools for collecting and storing data for the purpose of monitoring the machines condition. In addition, it is important to understand what data is required to make mobile from the overall company database, i.e. what type of data to gather and handle in order to make it mobile and access the mobile device. It is essential since some of the data is not appropriate to handle and process on the mobile device. For the reason that big amounts of data are needed for the decision making process and at the current stage, it is impossible to handle it on the mobile device because of constraints of the devices, such as hardware and processing power, especially when a CBM strategy is used.

The database development is normally carried out in two phases, i.e. the database design. These are the logical and physical database design phases. In logical design, the main activity is to identify the objects, the relationship between the objects, objects identifiers and classes. The physical database design concerns the development and application of the database based on the logical design. However, before the logical database design is done a conceptual database model should be designed [10]. This phase is independent of all implementation details in the same way as the underlying model, such as the relational or object data model design or other physical things which should be taken into account. Chen, [11] mentions that conceptual modelling used with the purpose for developing Information systems involves two

views, i.e. static and dynamic. When it comes to the static view it has to do with highlighting static properties, namely entities and relationships. While the dynamic view involves the way entities of the static design change over time, which is affected by the environment and in its turn triggers the changes [12]. Consequently, the conceptual modelling aims to build a representation of selected semantics about some real-world domain [13]. The conceptual model form bases on further design and programming phases in the system development for either building or modifying a software application system.

However, what is important to comprehend when developing e-maintenance application, especially for the mobile devices, is what kind of data of the whole enterprise to make accessible in the mobile devices. There are various projects which have defined ontologies for the mobile domain in the private sector, there are however no standards or agreement on a common semantics [14]. Similarly in the area of the e-maintenance the mobile ontologies or databases are none - existent. There are some efforts done in this area, such as Fugamalli et al. [15]. In addition, there are some researchers that have realized the importance of incorporating the e-learning aspects into the condition monitoring and maintenance systems [16; 17; 18; 19; 20; 21; 22]. However, in the current paper the authors consider the data and information needed for a mobile e-learning and knowledge sharing module when the maintenance engineer is mobile supported with such web technologies as the Web 2.0 social media technologies, i.e. the Wiki.

In the present work the authors present a conceptual database with the intention to use it with a mobile device. The conceptual database is suitable for a manufacturing or process plant and it has been developed in accordance with the recommended standards of the OSA-CBM (Open System Architecture for Condition Based Maintenance) and MIMOSA (Machinery Information Management Open Systems Alliance). Moreover, the e-learning module has been modelled to fit the needs of a maintenance engineer using a mobile e-maintenance system with the access to the required data when the user is on the move for purposes of efficient data processing based on the emergent web technologies, such as the Web 2.0, which might consist of the Wiki system when the user performs different work tasks in order to insert new information for its further successful access resulting in increased knowledge sharing and collaborative system. Further, the authors present the e-maintenance approach. Thereafter, in section 3 organisational knowledge and e-learning are mentioned. Next, the mobile database is presented in section 4. Finally, in section 5 the discussion part is given and in section 6 there are the conclusions.

2. E-MAINTENANCE

E-maintenance is quite a new term that is used to refer to the new prospects the development of ICT has facilitated with the aim of performing maintenance in the most effective way. In their review of e-maintenance related literature Levrat et al. [23] discuss the definition of e-maintenance which according to them differs depending on the authors. Thus, e-maintenance is a multi-faceted approach. It is more than implementing a maintenance strategy, maintenance plan or a maintenance type [24]. Through the framework proposed by Lung et al. [25] and Levrat et al. [17] it is possible to get an insight into the broad range of the e-maintenance concept. This is noted, especially, when observing the abstraction levels proposed by the authors. The abstraction levels are the e-maintenance strategic vision, e-maintenance busi-

ness 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. ICT, especially the web technologies, such as the Web Services, the semantic framework, etc., play a vital role in achieving this objective.

In this paper the authors have decided to take a slightly more straightforward approach where the key feature in e-maintenance is the huge improvement it gives in handling information. As for any decision, in this case for maintenance, the right information is crucial since the optimal time of maintenance is defined based on the information that the e-maintenance system provides for different users. When different work orders are created with the help of appropriate guidance in carrying out maintenance, next emerges the need to find the right machine that requires servicing. Also, a need arises in connection to the availability of, for instance, some spare parts. Figure 1 illustrates an overview of the questions and needs of a maintenance engineer. Figure 1 also shows a simplified summary of questions the maintenance engineer needs to find answers to.

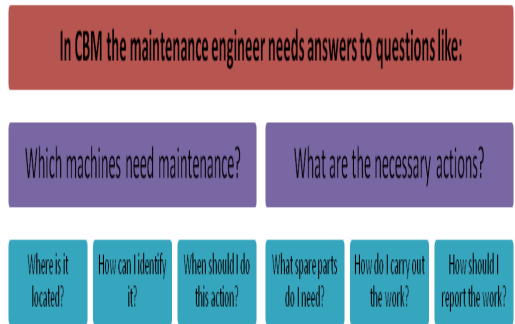


Figure 1. Example of questions a maintenance engineer needs to answer when following CBM strategy [26]

A comprehensive work within the e-maintenance based on the results of a European project DYNAMITE (Dynamic Decisions in Maintenance, IP017498) can be found in Holmberg et al. [27]. In the mentioned project a platform was developed, i.e. the DYNWeb, which is gone through in Holmberg et al. [27]. In addition, other efforts were done in the direction of the e-maintenance approach, such as TELMA e-maintenance platform, which is developed in Nancy University and is based on the CBM and proactive strategies Levrat and Lung [28]. Moreover, a recent work within the e-maintenance is the multi-layered intelligence platform named WelCOM for the optimal operation and maintenance of the equipment [29; 30]. Additionally, other work emphasizing the aspects of the e-maintenance approach can be found in Emmanouilidis et al. [31]. According to the above mentioned e-maintenance platforms enable the use of the web technologies, such as, the Web Services based on the semantic framework. Thus, such emerging technologies as the Web 2.0, i.e. the social media technologies, can be the tools to support the transition of the e-maintenance approach to incorporate knowledge driven approaches. Consequently, it can then support the use of knowledge man-

agement systems for purposes of knowledge sharing and e-learning within the e-maintenance approach. Consequently, in section three the organisational knowledge and e-learning are gone through.

3. ORGANISATIONAL KNOWLEDGE AND E-LEARNING

The value of organisational learning and knowledge is widely recognised and is as well viewed as a key asset that strengthens a company's competitive advantage. The e-learning can be seen as the ICT system that delivers the right information to the right person in the right amount, etc. In addition, the primary role of an e-learning system is to provide a collaborative environment that provides possibilities to create and maintain learning content. The access to information is essential for learning and instructions, therefore, the e-learning platforms play an important role in the e-learning process [32]. Thus, researchers argue that e-learning is the creation and distribution of the knowledge in an organization through the use of online delivery of information, communication, education, and training [33]. However, the e-learning approach and its applications are still in their infancy [34].

The mobile learning is a concept and research area that has emerged from the increasing availability of mobile devices. Mobile learning is born out of the concept of e-learning and is seen as a further development of e-learning as well as a subset of e-learning approach. The concept can be broadly described as learning that takes place using mobile devices. Authors that describe different aspects of the concept can be found in [35].

There are researchers that have tried to understand the characteristics of the e-learning, for instance, Sambrook [36] who explores the existing and potential role of e-learning in small and medium-sized organisations (SMEs). The author presents a model that identifies dimensions and factors that influence e-learning from both employee and employer's perspective. In addition, the author mentions that in the conducted survey it was shown that the barriers to implement e-learning in a small organisation were lack of hardware, lack of e-learning expertise, lack of time and resources as well as trust. Moreover, the difficulty in determining the full cost of e-learning initiatives was also an important factor that prevents the implementation of the e-learning approach in SMEs. Other factor of importance connected to the ICTs for the acceptance of the e-learning system was the level, i.e. that the material presented is of adequate level when it concerns knowledge and skills expected from learners. Moreover, other factors mentioned for the same purposes were user-friendly, clear and accurate presentation, adequate user interface, navigation, interesting material, language, etc.

Nevertheless, in the domain of interest different work tasks that the maintenance engineers normally perform are complex and differ depending on the fault occurred. Emmanouilidis et al. [31] mention that the education of the maintenance engineers in Europe is usually supported by bodies' recommendations CEN [37], or by National bodies, such as the Institute of Asset Management in the UK, IAM [38], or the European Federation of National Maintenance Societies EFNMS [39; 40]. The resources in an organisation and in this case maintenance are personified by the employee's experience, knowledge and skills, and it takes form in different working roles in the maintenance organisation.

Thus, it is important to aim at developing a knowledge sharing culture in the organisation where knowledge can be easily shared

with the support of the ICTs [41]. This is an important matter in the knowledge management literature, which emphasises the need of knowledge sharing culture, since it provides employees that are keen to share their knowledge with an increased collaboration. The employees' knowledge and learning as well as knowledge sharing are important issues for a successful organisation. The significance of knowledge is frequently overlooked in organisations. It is only when knowledgeable employees are lost, for instance, because of change of employment. This is relevant for any organisation, but even more valid for practice-oriented areas like maintenance. This can be explained due to the existence of two types of knowledge, i.e. tacit and explicit. The tacit knowledge is individual and is gained through personal experience and practical skills [42; 43]. While, the explicit is knowledge that can be stored, for instance, in a database and takes the form in instructions, procedures, etc. Explicit knowledge means knowledge that is clear, obvious and definite. Consequently, the practice-oriented areas like maintenance involve the two kinds of knowledge and the tacit is the one that is difficult to keep in the company because it is based on the employees' own experiences and practical skills, which are difficult to digitalise, transfer and share.

The difference between tacit and explicit knowledge advocate four elementary patterns to create knowledge in an organisation, i.e. socialization, externalisation, synthesis and internalization. The first, i.e. socialization, is "from tacit to tacit" and is when tacit knowledge is shared by one person at real time with another person. The second one means when tacit knowledge is transformed to new explicit knowledge, i.e. the delivery of best practices or lessons learned. The synthesis is referred to when explicit knowledge is used to create new knowledge. Internalization means when new tacit knowledge is created from explicit knowledge, i.e. the learning process that takes place from reading or discussion [42; 43]. Consequently, the objective of KM and KM systems is to support creation, transfer, and application of knowledge in organizations [44]. Thus, the KM approach highlights the need to share the information and knowledge in an organization, i.e. how information and knowledge in organizations can be organized and managed.

Such web technologies as the Web 2.0 have recently appeared. It's a concept and social media technology that are composed of a set of independent services, which provides a rich user with interaction and collaboration. In addition, this new technology, its applications and services, enables and facilitates the collaboration as well as information and knowledge sharing between employees in an organization.

The Web 2.0 and its social media web technologies become possible through the platform of Semantic Web, Web Services and the use of ontologies, etc. Examples of the Web 2.0, i.e. media technologies, are the wikis and blogs. However, it does not refer to new technical standards, but to new ways of using the Internet as a platform for interactive applications [45; 46; 47]. The Web 2.0, even named the social media technology, enables the capture, storing and sharing of not only information and knowledge (explicit knowledge), but also the tacit knowledge that a person or group has.. Crucial for successful implementation and use of the Web 2.0 technology is the organizational design, which needs to be seen as a socio-technical system. This is due to its dependence on the interaction between employers, organization and its technical systems for its greater effect. The web technologies, such as the Web 2.0 and the Semantic Web have come into various sectors, namely health and school sector, and lately there has been

noticed the emergence of the e-maintenance concept worth considering when developing modern maintenance applications that require knowledge sharing driven functionalities. The e-Maintenance achievements can reap substantial benefits from incorporating ICT tools and especially Web2.0 and social media technologies to support the organizational knowledge management process [31].

Next, different modules of databases are presented with the intention to be made accessible via a mobile device based on the web technologies, such as the Web Services and Web 2.0 technologies to support the maintenance engineer in different work tasks that are performed.

4. MOBILE DEVICE DATABASE

OSA-CBM (Open System Architecture for Condition Based Maintenance) and MIMOSA (Machinery Information Management Open Systems Alliance) have been working on standards for information exchange and communication among different modules for CBM, [48], (www.mimosa.org). MIMOSA developed a Common Relational Information Schema (CRIS). It is a relational database model for different data types that needs to be processed in a CBM application. The system interfaces are defined according to the database schema based on CRIS. The definitions of interfaces developed by MIMOSA are an open data exchange convention to use for data sharing in today's CBM systems. Defined by MIMOSA Cris is also MIMOSA'S Open System Architecture for Enterprise Application Integrating (OSA-EAI), which provides an open exchange standard for technology types in key asset management areas, such as asset register management, work management, diagnostic and prognostic assessment, vibration and sound data, oil, fluid and gas data, thermographic data and reliability information. Other important contribution in this area is the ISO 17359 standards, which suggest the reference values to use when a condition monitoring programme is started. In the figure below there are UML package, i.e. CMMS, and DSS and their subsystems, in this case for the CMMS the work order management and for the DSS the subsystems data acquisition, signals analysis, condition monitoring, diagnosis and prognosis. In the middle of Figure 2 there are the Web Services that fetch different data and information into the databases.

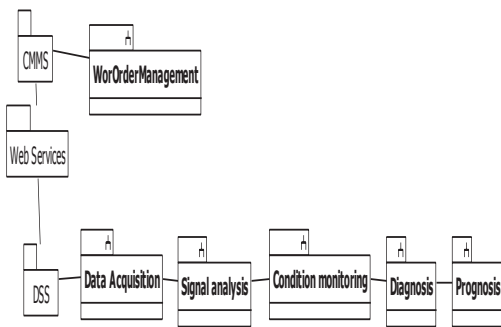


Figure 2. The e-CMMS and e-DSS [3]

Figure 3 shows on a conceptual level the MIMOSA Cris basic database classes needed for a CMMS handling, i.e. work management. Similarly, Figure 4 illustrates the minimum set of database classes to support the measurements of an asset. The boxes in these Figures represent the classes in the respective system. Inside each box there are the names of the classes specified. The boxes are connected through lines, which represent their relationship. The numbers shown between the classes specify the multiplicity (cardinality) in the specific relation. It specifies the number of classes/objects that can participate in a relation. In Figure 3, the relationship between WorkOrder and WorkOrderSteps is made through composition, which illustrates that WorkOrdersSteps and other classes connected to it are part of the whole, i.e. WorkOrder. While the relationship between the database with WorkAuditType in Figure 3 and TransducerAxisDirectionType in Figure 4 is done through aggregation, which is a rather loose connection.

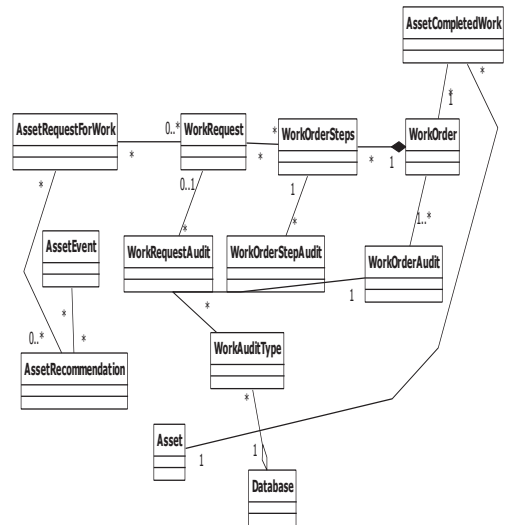


Figure 3. Work order database classes

The work management technology type allows the creation and auditing of a new work request in a work management system for a service segment or a serialized asset (www.mimosa.org). It allows as well the retrieval of work orders and the steps of work orders and the actual work completed information. For different measurements of an asset, MIMOSA provides modules named Diag and Dyn. The Diag technology enables the retrieval of human or "smart-agent" generated current and/or future proposed asset health states, current and/or future proposed diagnostic failure modes and casual trees, remaining useful life predictions, and recommendations. Also, it allows access to measurement evidence supporting the diagnoses/prognoses. These features are used by diagnostic and prognostic systems. The Dyn technology type enables the creation and retrieval of historical dynamic measurements (used with vibration and sound monitoring includ-

ing frequency spectra measurements and time waveforms), abnormal data alarms, and operational event logs.

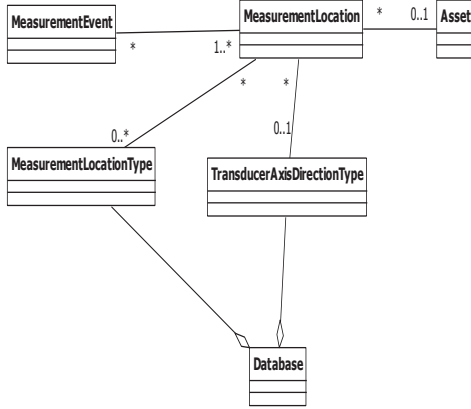


Figure 4. The measurement event database classes

A general conceptual database model has been developed, made suitable for a manufacturing or process plant and shown in Figure 5. In a manufacturing plant there are many machine-trains. Machines in each train then can be identified separately by a driving machine (an electric motor), an intermediate machine (a gear box) and a driven machine (for example, a pump). Maintenance decision, however, may not be taken in isolation. For the comprehensive maintenance decision-making, access to different databases of the enterprise is necessary.

The Condition Monitoring data may include parameters monitored, measured values of the monitored parameters, location of measurement, frequency of measurement, date and time of each measurement, reference values of the monitored parameters. The reference values may include allowable, just tolerable and not-permissible values. Often more than one condition monitoring parameter is monitored. For example, in case of a pump there may be vibration and performance parameters, such as flow rate, pressure generated and efficiency, etc. Vibration can be measured at different bearings (location) at an interval of a week (frequency) or even continuously depending upon the criticality of the machine. The temperature of the bearing (if it is a fluid film bearing) lubricant can be another parameter that should be monitored continuously. Depending on a particular condition monitoring parameter, there could be other details as well. For example, in case of vibration, the data may include also: velocity, acceleration or displacement, which is monitored as well as it is 0-peak, peak-peak, average or RMS values that are measured.

When developing the e-learning module the knowledge of the existent e-learning platforms was considered. It is important, as mentioned before, to consider the experience and advancement from, for instance, the distance learning sector in how the knowledge sharing can and should take place with the use of the Web 2.0 technologies, i.e. e-learning platforms, when developing

the e-maintenance applications that require knowledge sharing driven functionalities. Figure 6 illustrates the desirable classes since a mobile e-learning and knowledge sharing system requires providing the maintenance engineer with the suitable information and knowledge to make his work more efficient.

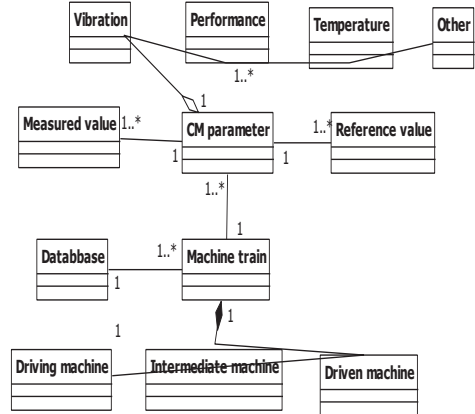


Figure 5. The CM Database Model

The user's role can be either the maintenance engineer or the expert. There can arise several Topics, i.e. parts of a CM and maintenance work task. Consequently, the Work order has a relation to the class Topic, which in turn has a relation to several Categories, i.e. different CM and maintenance activities. The Format to support the maintenance engineer can, for instance, be of several types, such as text, video, or picture, etc. In addition, a Wiki is connected to each Topic category as well as Format where it is assumed that the maintenance engineer shares his experience when performing different work tasks, which is accessible to other maintenance engineers later on.

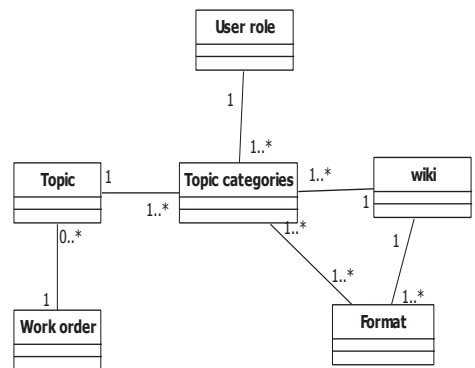


Figure 6. The e-learning module

All these classes and modules are later on connected to each other, i.e. Figure 3-6, into a complete database system where it is further on accessed with the Web Services and transferred into the client machines. In this case the mobile devices. The e-learning is supported by different classes, for instance, when the user gets knowledge transmitted through different formats, such as text, video and pictures, etc., in the right amount and to the right person. In addition, the Wiki provides possibilities to create and share knowledge between different users while carrying their mobile device resulting in an increased organisational knowledge as well as a collaborative environment. The different users' roles give the possibility for an expert to share/transmit his knowledge with/to other users' role, i.e. the maintenance engineer with the aim of increasing the efficiency of their different work task.

5. DISCUSSION

The introduction of e-learning in support with the latest web technologies, such as the Web 2.0 technologies, i.e. the Wiki systems, Web Services, and the use of the semantic web into the e-maintenance is important, since it brings the possibility to take advantage of the benefits of the knowledge management approach, such as the knowledge sharing between different employees in the domain of interest because of the characteristics of the suggested ICTs. Therefore, the use of the above mentioned web technologies and their Semantic web framework makes not only the knowledge sharing between employees possible but also the integration between software application systems and heterogeneous data, since, as already mentioned, the characteristics of these technologies facilitates and optimizes the aspects of the e-learning systems and knowledge sharing. Further, with the emergence of the Web 2.0 technologies bring new possibilities, such as the sharing knowledge between the employees through the Web 2.0 and its social media technologies as well as collaboration aspects. Consequently, it provides the maintenance engineer, using a mobile device for his different work tasks, with new opportunities. The above mentioned goes in hand with the statements of Peppard and Ward [49] and Pearson and Saunders [50], who state that organizations that are in pace with new developments of the ICTs ought to have a knowledge approach in their organization and use suitable ICTs to support them. This results in turn in higher organizational performance.

The possibilities that the above mentioned ICTs provide enable the maintenance engineer to receive the work order/s and the related information in different formats to acquire knowledge on how to proceed with the work task is crucial, since the variety of work tasks varies and their complexity as well. Accordingly, it increases the efficiency of the maintenance engineer when performing different work tasks. In addition, it also helps to avoid the situation when the knowledge is bound to some personnel which, in case when they quit the company for any reason, is lost, since this knowledge will be stored in a database in different formats, i.e. text, video, graphical, etc. As can be understood from the above mentioned, both the databases and the web technologies are important since without the data stored in the databases the ICTs tools are useless and without the web software applications the data is uncontrollable [47]. However, it is important to understand what kind of data is appropriate to make mobile since, for instance, the condition monitoring part of a Decision support system utilises huge amount of data when diagnosis is performed. It is, therefore, important that the data processing takes place at the

server and then later on the results are accessed via the mobile device.

The incorporation of the knowledge management and the e-learning approach with the support of the web technologies, such as the Web 2.0, into the e-maintenance becomes a reality because of the characteristics of these ICTs. The Web 2.0 facilitates the capture, storage and sharing of not only information and knowledge (explicit knowledge), but also the tacit knowledge that a person or group has-possesses. However, vital for the successful implementation and use of the Web 2.0 technologies is the organisational design for the reason that it needs to be seen as a socio-technical system, i.e. it depends on the interaction between employers, organization and its technical systems for its greater effect and successful implementation.

The maintenance decision, however, may not be taken in isolation. For the comprehensive maintenance decision-making, access to different databases of the enterprise is necessary. The Web Services agent can be utilized for the retrieving heterogeneous data from distributed databases over the Internet. It provides as well access to geographically distributed database through the Internet. Therefore, the authors believe that web technology, such as Web Services, could be a useful tool for fetching the data and information from geographically distributed databases for the purpose of e - monitoring and maintenance.

6. CONCLUSIONS

The paper proposes a conceptual database for a process/manufacturing plant taking into consideration the standards of OSA-CBM (Open System Architecture for Condition Based Maintenance) and MIMOSA (Machinery Information Management Open Systems Alliance). Also, the suggested database considers knowledge sharing and e-learning with the support of the Web 2.0 technologies, such as the Wiki, to be used in maintenance. It highlights important objects and classes, i.e. data and information that are required for practicing condition based and other maintenance strategies in a plant with the support of a knowledge sharing and e-learning module. In the context of global scenario where control of the machines condition, production, availability, performance, and quality is important, e-maintenance and e-monitoring are the tools that can enhance the enterprise maintenance effectiveness. In addition, the e-maintenance incorporating knowledge driven and e-learning approaches play a crucial role in solving the problem of non-availability of good experts that is a commonplace problem as well as facilitating a collaborative environment.

REFERENCES

- [1] Campos, J. (2009). Development in the application of ICT in condition monitoring and maintenance. *Computers in Industry* 60, 1–20. doi:10.1016/j.compind.2008.09.007
- [2] Campos, J., Jantunen, E., Prakash, O. (2007). Development of a Maintenance System Based on Web and Mobile Technologies. *Journal of International Technology and Information Management, JITIM*. 16. 1-8.
- [3] Campos, J., Jantunen, E., Prakash, O., 2009. A web and mobile device architecture for mobile e-maintenance. *Int J Adv Manuf Technol* 45, 71–80. doi:10.1007/s00170-009-1942-x
- [4] Wang, W., Tse, P.W., Lee, J., (2007). Remote machine maintenance system through Internet and mobile communi-

- cation. *Int J Adv Manuf Technol* 31, 783–789. doi:10.1007/s00170-005-0236-1
- [5] Jantunen, E., Giordamli, C., Adgar, A., Emmanouilidis, C., (2010). Mobile Devices and Services, in: Holmberg, K., Adgar, A., Arnaiz, A., Jantunen, E., Mascolo, J., Mekid, S. (Eds.), *E-Maintenance*. Springer London, pp. 227–246.
- [6] Irigaray, A.A., Gilabert, E., Jantunen, E., Adgar, A. (2009). Ubiquitous computing for dynamic condition-based maintenance. *Journal of Quality in Maintenance Engineering* 15, 151–166. doi:10.1108/13552510910961101
- [7] Campos, J., Jantunen, E., Prakash, O., (2009). A web and mobile device architecture for mobile e-maintenance. *Int J Adv Manuf Technol* 45, 71–80. doi:10.1007/s00170-009-1942-x
- [8] Antonelli, A., Boccaletti, C., Duni, G and Fabbri, G. (2007) Remote control and monitoring of an induction motor, in: *Proceedings of COMADEM, the 20th International Congress of Condition Monitoring and Diagnostic Engineering Management*, pp. 505 – 513.
- [9] Kanda, S., K. Senba., y. Nanakaya., H. Ikeda., T. Kawai (2000). "Database management system for proactive maintenance (case study of steam turbine plant and switch-house equipment in Japan)", 2000 IEEE Power Engineering Society Winter Meeting. Conference Proceedings, Vol. 1, pp. 464-9.
- [10] Connolly, T. and C.E. Begg (2004). *Database Solutions: A step by step guide to building databases*, 2nd ed. Addison-Wesley, pp 191 - 193.
- [11] Chen, P. (1976). The entity-relationship model: toward a unified view of data. *ACM Transactions on Database Systems* Vol. 1, No. 1, pp-9–36.
- [12] Booch, G, Rumbaugh, J and Jacobson, I. (1999). *The Unified Modelling Language User Guide*. Addison-Wesley, Harlow, UK.
- [13] Weber, R., (2003). *Conceptual Modelling and Ontology: Possibilities and Pitfalls*, in: Spaccapietra, S., March, S.T., Kambayashi, Y. (Eds.), *Conceptual Modeling — ER 2002*, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 1–2.
- [14] Banu, W.A., Khader, P.S.A., Shriram, R., 2011. *MOBILE ONTOLOGY DESIGN FOR SEMANTIC WEB: A CASE STUDY*.
- [15] Fumagalli, L., Jantunen, E and Macchi, M. (2009). *Ontology of Mobile Maintenance Processes*, proceedings of the 2009 Conference of the Society for Machinery Failure Prevention Technology, MFPT 2009, Dayton, Ohio, USA, April 28-30, 2009, pp. 369-386.
- [16] Nof, S.Y., 2003. Design of effective e-Work: review of models, tools, and emerging challenges. *Production Planning & Control* 14, 681–703. doi:10.1080/09537280310001647832
- [17] E. Levrat and B. lung. "TELMA: A full e-maintenance platform" in *Centre de Recherche en Automatique de Nancy Université*, 2007
- [18] Morel, G., Valckenaers, P., Faure, J.-M., Pereira, C.E., Die-drich, C., 2007. Manufacturing plant control challenges and issues. *Control Engineering Practice*, Special Issue on Manufacturing Plant Control: Challenges and Issues INCOM 2004 11th IFAC INCOM'04 Symposium on Information Control Problems in Manufacturing 15, 1321–1331. doi:10.1016/j.conengprac.2007.05.005
- [19] Jantunen, E., Gilabert, E., Emmanouilidis, C., Adgar, A., 2010. e-Maintenance: a means to high overall efficiency, in: Kiritisis, D., Emmanouilidis, C., Koronios, A., Mathew, J. (Eds.), *Engineering Asset Lifecycle Management*. Springer London, pp. 688–696.
- [20] Emmanouilidis, C., Spais, V., 2010. E-training in Maintenance, in: Holmberg, K., Adgar, A., Arnaiz, A., Jantunen, E., Mascolo, J., Mekid, S. (Eds.), *E-Maintenance*. Springer London, pp. 475–506.
- [21] Jantunen, E., Adgar, A., Emmanouilidis, C & Arnaiz, A. (2010), "Next generation maintenance through the adoption of e-maintenance", *Proceedings of the original from Proceedings of the Euromaintenance 2010 Conference, 12-14 May 2010, Verona, Italy (EFNMS)*.
- [22] Papathanasiou, N., Emmanouilidis, C., Pistofidis, P., Karampatzakis, D., 2012. E-Learning and Context Aware e-Support Software for Maintenance. *J. Phys.: Conf. Ser.* 364, 012110. doi:10.1088/1742-6596/364/1/012110
- [23] Levrat E., lung B. & Crespo Marquez A. (2008). e-Maintenance: review and conceptual framework. *Production Planning & Control*, 19 (4), pp 408-429.
- [24] Crespo-Márquez A. (2008). *The Maintenance Management Framework: Models and Methods for Complex Systems Maintenance*, Springer Series in Reliability Engineering.
- [25] lung, B., Levrat, E., Marquez, A.C., Erbe, H., 2009. Conceptual framework for e-Maintenance: Illustration by e-Maintenance technologies and platforms. *Annual Reviews in Control* 33, 220–229. doi:10.1016/j.arcontrol.2009.05.005
- [26] Jantunen E., Macchi M., Fumagalli L. & Di Leone E. (2010). *Financial Analysis of e-Maintenance, Maintenance & Asset Management Maint World*, No. 2, pp 38-43.
- [27] Holmberg K., Adgar A., Arnaiz A., Jantunen E., Mascolo J. and Mekid S. (2010). *E-maintenance*, Springer, ISBN 978-1-84996-204-9.
- [28] Levrat, E. and lung, B. (2007), "TELMA: A full e-maintenance platform", *Proceedings of WCEAM/CM, 2nd World congress on Engineering Asset Management, Harrogate UK*.
- [29] Papathanasiou N., Emmanouilidis C., Pistofidis P and Karampatzakis D. (2012), "E-Learning and Context Aware e-Support Software for Maintenance", 25th International Congress on Condition Monitoring and Diagnostic Engineering.
- [30] Pistofidis, P and Emmanouilidis, C. (2012), "Developing Advanced Context Aware Tools for Mobile Maintenance", *Proceedings of the IFAC-A-MEST 2012 Workshop on Advanced Maintenance Engineering*, (pre-print of the original published by IFAC).
- [31] Emmanouilidis, C., Jantunen, E., Gilabert., and Starr, A. (2011). *e-Maintenance update: the road to success for modern industry*. *Proceedings of the 24th International Congress on Condition Monitoring and Diagnostics Engineering Management, COMADEM 2011*.

- [32] Sanderson, P.E. (2002). E-learning: strategies for delivering knowledge in the digital age, *Internet and Higher Education*, Vol. 5, pp. 185–188.
- [33] Wild, R.H., Griggs, K.A & Downing, T.A. (2002). A framework for e-learning as a tool for knowledge management, *Industrial Management & Data Systems*, Vol. 102, Issue 7, pp.371 – 380.
- [34] Kahiigi, E.K. ., Ekenberg, L. ., Hansson, H. ., Tusubira, F.F & Danielson, M. (2008). Exploring the e-Learning State of Art.” *The Electronic Journal of e-Learning* Vol. 6, Issue 2, pp.77 – 88.
- [35] Zuga,B., Slaidins, I., Kapenieks, A., & Strazds, A. (2006). M-learning and Mobile Knowledge Management: Similarities and Differences, *International Journal of Computing & Information Sciences* Vol. 4, No. 2, pp. 58 – 62.
- [36] Sambrook, S. (2003). E-learning in small organisations, *Education + Training*, Vol. 45, Issue 8/9, pp. 506-516.
- [37] CEN 2007, CEN/TR 15628:2007 Maintenance. Qualification of maintenance personnel.
- [38] IAM 2008, IAM PAS55:2008 competence framework. Institute of Asset Management, London.
- [39] EFNMS 1998. The Requirements and Rules to achieve an EFNMS Certificate as a European Expert in Maintenance Management.
- [40] EFNMS 2000. The specification for requirements of a European Maintenance Technician Specialist.
- [41] Jashapara, A. (2004) *Knowledge Management: An Integrated Approach*. England: Pearson Education Limited.
- [42] Nonaka, I., Ichijo, K. & Krogh, G., (2000). *Enabling Knowledge Creation: how to unlock the mystery of tacit knowledge and release the power of innovation*. Oxford University Press, New York
- [43] Rowley, J & Farrow, J. (2000). *Organizing knowledge-An introduction to managing access to information*, 3 edition Ashgate Publishing Limited
- [44] Alavi, M., Leidner, D.E., 2001. Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly* 25, 107–136. doi:10.2307/3250961
- [45] Thies, G. and Vossen, G. (2008). “Web-oriented architectures: On the impact of web 2.0 on service-oriented architectures,” *Asia-Pacific Conference on Services Computing*. 2006 IEEE., pp. 1075–1082.
- [46] Venkatraman, S.S. (2004). Web-Services – The Next Evolutionary Stage Of E-Business, *Journal of International Technology and Information Management*, Vol.13, no 2, pp. 111-122.
- [47] O’Reilly, T. (2005). *What is Web 2.0—Design Patterns and Business Models for the Next Generation of Software*, Available at <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-isweb-20.html>.
- [48] Thurston, M.G., 2001. An open standard for Web-based condition-based maintenance systems, in: *AUTOTESTCON Proceedings*, 2001. IEEE Systems Readiness Technology Conference. Presented at the AUTOTESTCON Proceedings, 2001. IEEE Systems Readiness Technology Conference, pp. 401–415. doi:10.1109/AUTEST.2001.949021.
- [49] Peppard, J., Ward, J., 2004. Beyond strategic information systems: toward an IS capability. *Strategic Information Systems* 13, pp. 167–194.
- [50] Pearlson, K. E. & Saunders, C. S. (2006). *Managing and Using information systems – A strategic Approach*, 3rd edition, Wiley.

Effect of the shape of connecting pipes on the performance output of a closed-loop hot water solar Thermo-syphon

B. Freegah

University of Huddersfield
Huddersfield, UK
HD1 3DH
basim.freegah@hud.ac.uk
Al-Mustansiriya University,
Baghdad, Iraq

T. Asim

University of Huddersfield
Huddersfield, UK
HD1 3DH
t.asim@hud.ac.uk

D. Albarzenji

University of Huddersfield
Huddersfield, UK
HD1 3DH
dlir.albarzenji@hud.ac.uk

S. Pradhan

University of Huddersfield
Huddersfield, UK
HD1 3DH
s.r.pradhan@hud.ac.uk

R. Mishra

University of Huddersfield
Huddersfield, UK
HD1 3DH
r.mishra@hud.ac.uk

ABSTRACT

In order to conserve the environment from pollution, which is caused by the use of the fossil fuels, numerous research works have been carried out in renewable energy area to minimize the dependency on the fossil fuels. There are several energy sources naturally available, and solar energy is considered to be the best amongst them. Therefore it became a motivating area for the researchers in recent years. Thermo-syphon is one of many devices that use solar energy for power generation. Thermo-syphon converts solar energy into internal energy of the working fluid; mainly water. In this work, a computational fluid dynamics (CFD) code has been used to analyse the natural convection phenomenon in a thermo-syphon. The thermo-syphon model consist of steel pipes with an internal diameter of 25mm, along with a condenser having diameter equal to five times the pipe's diameter, has been considered. The study has been carried out under no-loading conditions, for two thermo-syphon models comprising of straight and helical shaped pipes of 10, 20 and 30. A practical solar heat flux of 500W/m² has been applied on the pipes. The numerical results depict that the working fluid within the condenser, in case of helical pipes, gains higher temperature as compared to the straight pipes. Furthermore, increase in the number of helical pipes has negligibly small effect on the temperature of the fluid within the condenser, and hence on the performance output of the thermo-syphon.

Keywords

Computation Fluid Dynamics (CFD), Thermo-syphon, Helical pipe, Natural Convection

NOMENCLATURE

q	Heat flux (W/m ²)
U	Overall heat transfer coefficient (W/m ² . °C)
ΔT	the difference in temperature (C°)
T	temperature of water within the condenser (°C)
T_{avg}	average temperature within the condenser (°C)
t	time of operation (minute)
t_{avg}	average time (minute)
nt	number of turns
np	number of pipes

1. INTRODUCTION

The amount of heat transfer through any material depends on several parameters; such as temperature variations, overall heat transfer coefficient, heat transfer surface area etc., as shown in Eq. (1).

$$q = U\Delta T \quad (1)$$

In recent decades, many researchers have been trying to improve the design of thermo-syphon in order to obtain excessive amount of useful heat energy. KE Amori et al. 2012 conducted a comparative study between a traditional absorber and a new design of solar collectors (known as the accelerated absorber) to analyse the performance of these systems. The performance evaluation was carried out in identical conditions for both the systems, having tilt angle of 33°. The evaluation tests were carried out in the presence of two different types of storage tanks. The results have shown that there is a significant increase in the thermal performance of the thermo-syphon (approximately 60%) for the new system. It has been observed that the temperature within the storage tank for the new design is 13C° higher to the conventional type. Subramanian et al. 2012 studied the impact of riser arrangement (zigzag pattern) on the performance of a flat plate collector system, and compared it with the conventional system. Experiments were conducted using copper tubes in header and riser having various geometrical characteristics. The results have shown that the performance efficiency reached 62.9% in the zigzag arrangement. El-Din et al. 2005 experimentally investigated the properties evaluation of the heat transfer in single-phase flow. In their study, a toroidal thermo-syphon type has been used. The parameters of investigation include heated-cooled length ratio, heated length tube diameter ratio, diameter ratio of torus-tube, and angle of inclination. Their results show that the increase in both heated-cooled length ratio and heated length-tube diameter ratio leads to decrease in the heat transfer rate, whereas increase in torus-tube diameter ratio increases the heat transfer rate. Furthermore, it was found out that the range of tilt angles between 30° and 45° produces maximum heat transfer rate. Freegah et al 2013 numerically studied the effects of the length to diameter ratio of the riser, number of connecting pipes, angle of inclination of the thermo-syphon and the heat flux, on the performance of the

thermo-syphon. It was found that the heat flux and the length to diameter ratio of the pipes have significant effects on the performance of a thermo-syphon, whereas, the angle of inclination has negligibly small effect. Furthermore, an increase in the number of connecting pipes increases the temperature of the working fluid, as they absorb more solar energy. Gurveer et al 2014 conducted an experimental study to investigate the effects of the inclination angle, wire coil inserts and wire mesh inserts on the thermal performance of a flat-plate solar collector. The results that have been reported indicate that the Nusselt number in novel insert configuration is higher as compared to the conventional system, and hence the thermal performance for the novel insert configuration has been observed to be better than the conventional one.

This study is the continuation of Freegah et al. 2013 and hence the natural convection phenomena, and the distribution of temperature and velocity of the working fluid, has not been discussed in detail in the present study. Computational Fluid Dynamics based tools have been used to carry out an extensive numerical study, on the effects of using helical pipes, on the performance of a closed-loop solar hot water thermo-syphon system. The effects of helical pipes in a thermo-syphon have not been explicitly analysed in the literature, and hence this study is important for the design process of such systems.

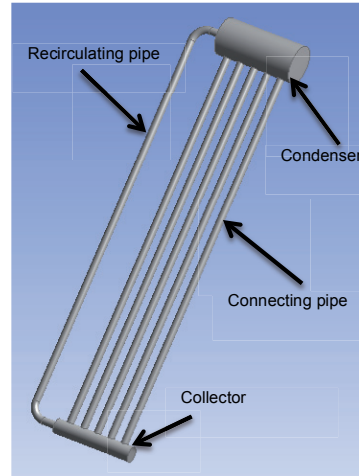
2. NUMERICAL MODELLING

Two thermo-syphon configurations, comprising of helical and straight pipes, have been modelled. The geometry of the two models is shown in Figure 1. An internal diameter of 25mm has been used for both helical and straight connecting pipes, with a thickness of 2mm. Furthermore, the recirculating pipe, which has been used in both the models, has the same diameter and thickness as that of the connecting pipes. It has been assumed that the thermo-syphon is operating under no-load condition. The diameter of the condenser is five times the diameter of the pipes, and the diameter of the collector is twice as that of the pipes. The working fluid considered is water.

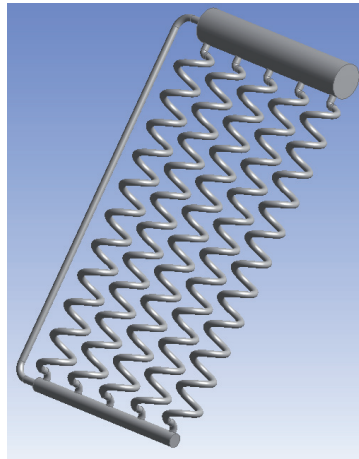
Hybrid meshing has been employed, using both hexagonal and tetrahedral elements. Non-uniform mesh distribution has been used, where the mesh elements are concentrated near the wall region, using 10 layers of mesh elements. The mesh contains two million elements, and has been shown previously to describe the flow phenomena with reasonable accuracy. Furthermore, a time step size of 12sec has been used.

Boussinesq approximation has been employed to accurately model buoyant forces being generated. This approximation states that the density differences are sufficiently small to be neglected, except where they appear in terms multiplied by g i.e. the acceleration due to gravity. The essence of the Boussinesq approximation is that the difference in inertia is negligible, but gravity is sufficiently strong to make the specific weight appreciably different. Furthermore, it has been observed by Dehdakel et. al. 2010 that the Boussinesq approach for the density of the working fluid in a thermo-syphon gives fairly accurate results, and thus has been used in the present study.

Three dimensional Navier-Stokes equations, in addition to the continuity and the energy equations, have been numerically solved in an iterative manner to simulate the transient flow of water in the thermo-syphon for one hour of operational time.



(a)



(b)

Figure 1. The geometry of the thermo-syphon (a) straight pipe and (b) helical pipe model

3. Results and Analysis

Freegah et al 2013 successfully simulated the natural convection phenomena in thermo-syphon; the simulations conducted to investigate the effect of number of the straight pipes, the tilt angle and the length-to-diameter ratio on thermo-syphon's performance. Authors presented numerical results in the form of temperature contour within the thermo-syphon to show the natural convection phenomenon (figure 2).

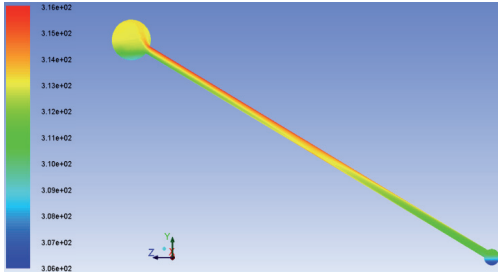


Figure 2. Temperature variations in the working fluid in the thermo-syphon after 1 hour of operation

In the present study, which is a continuation of Freegah et al 2013, the effect of the number of turns in a helical pipe, on the temperature within the condenser, has been numerical analysed. The boundary conditions are the same for all the numerical simulations.

3.1. Straight pipes

Figure 3 depicts the temperature distribution within the cross-section of the condenser of the thermo-syphon model comprising of straight connecting pipes. It can be clearly seen that the hot working fluid occupies the upper section of the condenser while the cold working fluid settles on the bottom of the condenser.

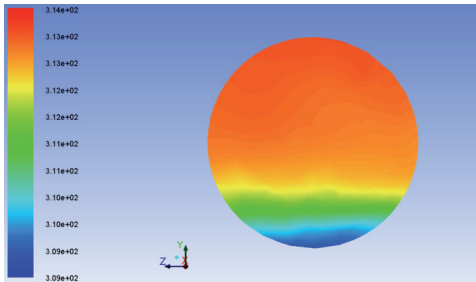


Figure 3. Temperature variations in the condenser after 1 hour of operation

Figure 4 depicts the variation in working fluid's temperature within the cross-section of the condenser for variable number of the connecting pipes. It can be seen that the temperature within the condenser is slightly higher for five connecting pipes as compared to three connecting pipes. It is obvious from the fact that more number of connecting pipes transfers more hot fluid to the

condenser, increasing its temperature. After one hour of operation, the difference in the condenser's temperature is 20C for both thermo-syphon configurations.

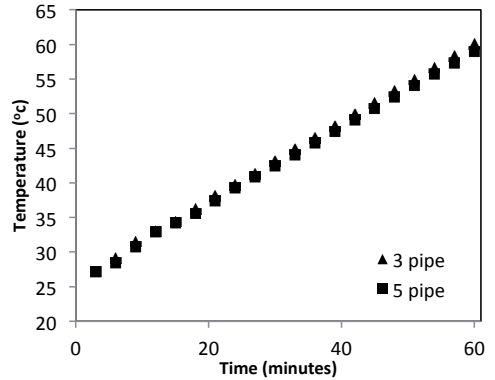


Figure 4. Variations in temperature of the working fluid within the condenser for various numbers of straight pipes

3.2. Helical pipes

Figure 5 shows comparison of the temperature within the condenser for the straight and helical connecting pipes comprising of 10 turns. It can be clearly seen that the hot working fluid occupies the upper section of the condenser while the cold working fluid settles on the bottom of the condenser. It can also be seen that the average temperature of water is higher for helical pipe as compared to the straight pipe. This shows that by using helical pipes, the temperature within the condenser can be increased.

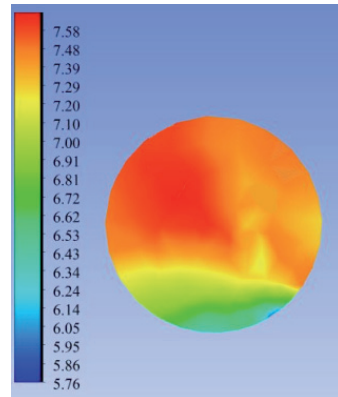
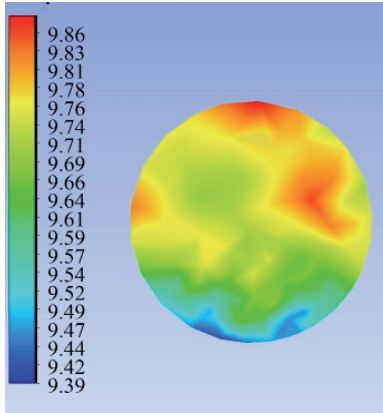


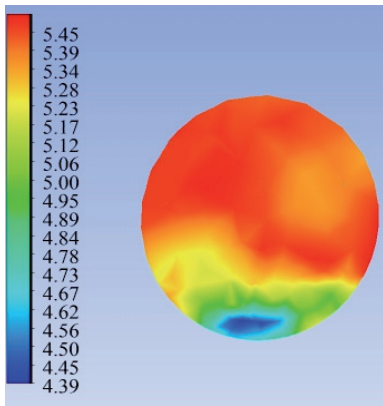
Figure 5. Comparison in condenser's temperature between straight pipe and helical pipe models consisting of 10 turns

Figure 6 depicts the effect of the number of turns of the helical pipes on the condenser's temperature after one hour of operating

time. Figure 2(a) shows the difference between 20 and 10 turns, while figure 2(b) shows the difference between 30 and 20 turns. It can be seen that the temperature of the working fluid increases as the number of turns increases. This is true for both the cases i.e. increase from 10 turns to 20 turns, and increase from 20 turns to 30 turns. After one hour of operation, temperature difference within the condenser for straight pipes, 10 turns, 20 turns and 30 turns are 9.86°C, 9.3°C, 5.4°C, 4.39°C respectively.



(a)



(b)

Figure 6. Comparison in condenser's temperature between (a) 20 and 10 turns (b) 30 and 20 turns

Figure 7 depicts the variations in the condenser's temperature for three helical models. It can be seen that the temperature of the condenser increases linearly in all the different cases. Furthermore, the condenser's temperature is considerably higher in case of 30 turns as compared to 20 and 10 turns' cases, after one hour of

operation. Meanwhile, the difference in the condenser's temperature between 10 turns and 20 turns is significantly higher than the difference between 20 turns and 30 turns. This is because the increase in the volume of the working fluid from 10 turns to 20 turns model is 100%, whereas it is 50% from 20 turns to 30 turns.

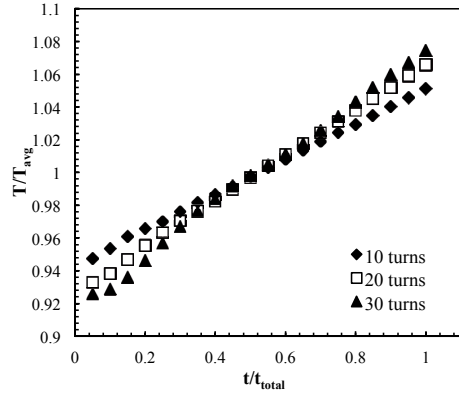


Figure 7. Variations in condenser's temperature for three models consisting of 10, 20, and 30 turns

Figure 8 depicts the variations in the condenser's temperature with 10 turns configuration, however with different number of helical pipes (3 and 5 connecting pipes). After one hour of operating time, the difference in condenser's temperature is 1°C between 3 and 5 connecting pipes. It can be seen that the working fluid's temperature within the condenser is negligibly higher for 5 connecting pipes as compared to 3 connecting pipes.

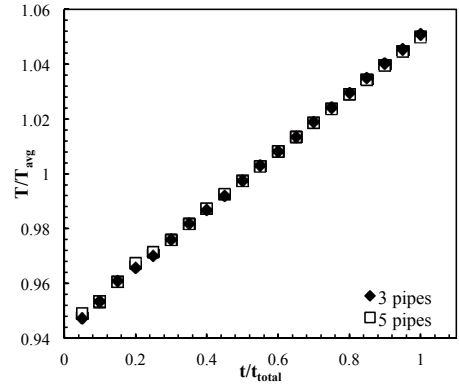


Figure 8. Variations in condenser's temperature for two helical models consisting of 3 and 5 connecting pipes

Temperature variations in the working fluid within the condenser for straight and helical pipes, at different times of operation, have been summarised in table 1. It can be clearly see that increase in

time of operation and number of turns lead to increased temperature within the condenser. For example, for three connecting pipes, after one hour of operation, the temperature of the working fluid increases by 12.66%, 24.78%, and 30.10% for 10, 20 and 30 turns respectively, compared to straight pipes model.

Table 1. Temperature of water in straight and helical pipe models at different times of operation

Time (minute)		15	30	45	60
Temperature in 3 straight pipes model (C°)		33.40	39.80	46.10	52.40
Temperature in 5 straight pipes model (C°)		34.81	42.54	50.19	57.79
Temperature in 3 helical pipes model (C°)	10 turns	34.40	43.10	51.60	60.10
	20 turns	36.80	47.50	58.60	69.80
	30 turns	37.10	50.50	62.10	75.10
Temperature in 5 helical pipes model (C°)	10 turns	34.20	42.20	50.71	58.99
	20 turns	36.06	46.53	57.67	68.54
	30 turns	35.29	50.16	60.51	74.13
	20 turns	3.40	8.50	12.90	15.67
	30 turns	1.35	15.18	17.10	22.03

From the numerical results, using multiple regression analysis, equation (2) has been formulated in order to calculate working fluid's temperature within the condenser for various thermo-syphon configurations. In order to check validity of Eq. (2) T/T_{avg} values have been calculated using Eq. (2), and compared against the results presented in table 1. It has been observed that Eq. (2) gives an average percentage error of 9.3% compared to the results presented in table 1, and hence this equation can be used to predict the temperature within the condenser with 90.7% accuracy.

$$\frac{T}{T_{ave}} = \begin{cases} \frac{(10^{0.028}) \left(\frac{t}{t_{avg}}\right)^{0.454}}{(nt)^{0.0124} (np)^{0.00137}} & \text{for } 0 < \left(\frac{t}{t_{avg}}\right) \leq 0.8 \\ (10^{0.028}) \left(\frac{t}{t_{avg}}\right)^{0.454} (nt)^{0.0124} (np)^{0.00137} & \text{for } 0.8 < \left(\frac{t}{t_{avg}}\right) \leq 1.6 \end{cases} \quad (2)$$

4. Conclusions

In the present work, CFD simulations have been conducted for thermal performance evaluation for two types of thermo-syphon, namely helical connecting pipes and conventional thermo-syphon. From the numerical results, it can be concluded that a considerable enhancement in the performance output of the thermo-syphon is obtained for the helical pipe configurations, in comparison with the conventional model. Furthermore, increasing the number of turns in the helical connecting pipes increases the condenser's temperature. Moreover, increasing the number of helical connecting pipes does not enhance the performance of the thermo-syphon significantly. It is expected that this study will help in the design process of thermo-syphons with optimal thermal performance.

5. References

1. B. Freegah, T. Asim and R. Mishra (2013). "Computational Fluid Dynamics based Analysis of a Closed Thermo-syphon Hot Water Solar System", 26th International Congress of Condition Monitoring and Diagnostic Engineering Management, Helsinki, Finland.
2. Dehdakel, A. A. Rahimi, M. and Alsairafi, A. A. (2010). "CFD modeling of flow and heat transfer in a thermo-syphon", International Communications in Heat and Mass Transfer, Vol. 37, pp. 312-318.
3. G. Sandhu, K. Siddiqui and A. Garcia (2014). "Experimental study on the combined effects of inclination angle and insert devices on the performance of a flat-plate solar collector", International Journal of Heat and Mass Transfer 71, pp. 251–263.
4. J. Subramanian, S. Tamilkozhndu and T. Selvam (2012). "Experimental Studies on Variable Header Solar Water Heating System", 2nd International Conference on Mechanical, Production and Automobile, 2nd International Conference on Mechanical, Production and Automobile Engineering Singapore April 28-29.
5. K.E. Amori and N.S. Jabouri (2012). "Thermal Performance of Solar Hot Water Systems Using a Flat Plate Collector of Accelerated Risers", TJER, Vol. 9, No. 1, pp. 1-10.
6. S. El-Din, H. S. (2005). "Heat transfer characteristics of an inclined single-phase toroidal thermo-syphon", Alexandria Engineering Journal, 44(2), 157-171.

eMaintenance 4A

Safety & Security in eMaintenance: The need for integration

Tim Tinney
Saab AB
Support and
Services
581 88 Linköping
+46 (0)734371726
timothy.tinney@
saabgroup.com

Anders Adlemo
School of
Engineering
Jönköping University
555 11 Jönköping
+46 (0)36101606
anders.adlemo@
jth.hj.se

Ulf Seigerroth
School of
Engineering
Jönköping University
555 11 Jönköping
+46 (0)36101785
ulf.seigerroth@
jth.hj.se

Olov Candell
Saab AB
Support and
Services
581 88 Linköping
+46 (0)734181739
olov.candell@
saabgroup.com

ABSTRACT

With the advent of eMaintenance, there have been many new possibilities and opportunities to increase the productivity of industrial systems, yet decrease resources and administrative costs. To accomplish this though, it is often required to rely on some network connectivity with the system of interest. It is this network connectivity capability that runs the risk of being interdicted, thereby providing a window of opportunity, commonly referred to as vulnerability, for a threat to exploit. This exploitation can occur in the form of a network security threat, information technology (IT) security threat, or information assurance and security threat. This in itself is not necessarily new, but the applicability to eMaintenance is not well understood. To gain a better understanding of this significance, there potentially needs to be a better conceptualization of how to integrate safety and security at an analytical level focused on the product or system design requirements. To accomplish this, the first question that needs to be addressed is whether or not it is practical, or even useful, to pursue the idea of integrating safety and security for the support and protection of eMaintenance solutions? If the answer to this is yes, then what type of integration would be recommended? If the answer to the first question is no, then why?

There have been a number of comparisons and contrasts between safety and security over the last couple of decades. In the end though there still has not been a viable proposed conceptualization that can be utilized to integrate different perspectives and practices in the safety and security domains. Today the most likely, although not the only, method to account for the integration of both domains is to conduct a safety analysis of a product or system of interest first and then afterward, conduct a security analysis, or vice versa. This method of development leaves much to be desired, and does not consider the potential overlap of similarities between safety and security. Nor does this method take into consideration that the inherent differences between the safety and security domains are significant to the preservation of the original analytical strengths of both domains.

Besides addressing the question of the practicality and usefulness of integrating safety and security, this paper will also address past research and noteworthy projects that have previously attempted this endeavour. Some general safety and security aspects of eMaintenance will also be addressed.

Keywords

eMaintenance, harmonization, information communication technology (ICT), integration, network connectivity, system safety, system security, unification, threat, vulnerability.

1. INTRODUCTION

With the inception of eMaintenance as a concept and field of research since the early 2000s, there have been many new possibilities and opportunities to increase the productivity of industrial systems, yet decrease required resources and administrative costs. One of the consequences of this eMaintenance rapid growth and development comes from the utilization of new technologies that can often potentially provide a ripe and fertile environment for “others” with a less than ethical perspective to take advantage of this situation, e.g. intellectual property infringement. This is not a line of logic that most commercial managers or leaders want to discuss or ponder.

To complicate matters, eMaintenance technologies are not getting any simpler. What is needed immediately is a way to better understand the environment that these existing and new eMaintenance technologies are being utilized through the use of more effective and viable conceptualizations.

To this end, there is no one single definition of eMaintenance that has been considered dominant. This could be for many various reasons, but giving the many applicable domains that eMaintenance can be utilized in today, this should not be surprising. Although, there have been a number of proposed eMaintenance definitions, some of which are as follows (underlining has been by the authors of this paper:

- eMaintenance is defined as the part of maintenance support that ensures that the maintenance process is aligned with the operation and modification processes to obtain business objectives, through proper information logistics by Information and Communication Technology (ICT) utilization and provision of information services [23].
- eMaintenance is seen as the application of ICT for remote maintenance and the representation of the physical world in a digital mode that aims at supplying tailored information as decision-support regarding appropriate maintenance activities for all stakeholders independent of time, geographical location, or organizational belonging [24].
- Maintenance support which includes the resources, services and management necessary to enable proactive decision process execution. This support includes e-technologies (i.e. ICT, web-based, tether free, wireless, infotonics technologies) but also eMaintenance activities (operations and processes) such as eMonitoring, eDiagnosis, ePrognosis [34].

- eMaintenance is a multidisciplinary domain based on maintenance and information and communication technologies (ICT) ensuring that the eMaintenance services are aligned with

the needs and business objectives of both customers and suppliers during the whole product lifecycle [22].

- eMaintenance is the application of Information and Communication Technology (ICT) for remote and online operation and maintenance activities for providing decision support to operation and maintenance process for all stakeholders independent of time, geographical location, or organizational belonging [20].

- eMaintenance can be defined as a maintenance strategy where tasks are managed electronically using real time equipment data obtained through digital technologies (i.e. mobile devices, remote sensing, condition monitoring, knowledge engineering, telecommunications and Internet technologies [46].

The aforementioned definitions are but a few of the definitions of eMaintenance that have been presented over the years. Other definitions can be found in literature but most of today's researchers seem to agree that in order for an eMaintenance definition to hold, it must contain some form of computing capability and a communication capability, for instance the inclusion of information communication technology (ICT). The number of definitions substantiates the need for ICT.

Another purpose of the above definitions is to emphasize the role of eMaintenance relative to traditional maintenance, and the importance of ICT within eMaintenance. From the definitions in Table 1, it can be inferred that maintenance is in support of operations, and that eMaintenance is support of maintenance operations. To accomplish eMaintenance requirements, it is of the utmost importance to possess viable and functional ICT services and capabilities. Thus, the conclusion is that ICT services and capabilities play a significant role in operations, as well as being vital to any operation using eMaintenance.

Candell [11] states that the importance of ICT in eMaintenance is as follows:

ICT is one of the main prerequisites, not only to improve the effectiveness and efficiency of the maintenance process for complex systems with long life cycles, but also to reduce associated risks, and contribute to a more efficient business process. The utilization of ICT offers a more controlled content sharing, information exchange and knowledge management within the phases of the maintenance process, coordination of the maintenance process with other processes (e.g. operation and modification processes) and a connection to strategic business objectives and external stakeholder requirements.

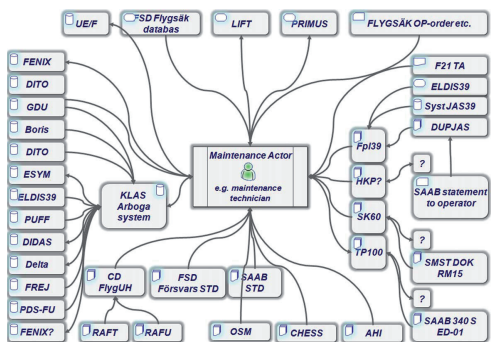


Figure 1. Information system needs of a Saab JAS 39 Gripen maintenance technician or planner [11]

One reason for utilizing ICT in eMaintenance has been that many of the current systems have grown so complex that they have become virtually impossible for one single person, or even a group of persons, to operate and maintain without relying on the help of some form of an ICT enabled system [11]. An example of this is shown in Figure 1.

Figure 1 also illustrates the complexity of the requirements and activities involved in the maintenance of one of today's fighter aircraft. With support from ICT, the performance of the maintenance activities can be significantly improved.

As it is today, the focus within eMaintenance has been more on the advantages and seemingly leaving the disadvantages to be solved at a later time. This is pertinent today because of the predominate focus on incorporating new technologies such as increased usage of wireless network technologies [29] to interconnect eMaintenance sensors [35] and provide for oversight and control center capabilities. The control center is where the eMaintenance data can be analyzed statistically [47] and consequently reacted upon. An example of this is through the use of intelligent prognostics tools [28]. Finally, all of the eMaintenance activities are being performed in real-time, thus the demands on the ICT enabled system are even more demanding.

A problem though is that most technological solutions come with both advantages and disadvantages. This is also the case with ICT in eMaintenance, more specifically. Being that there are many reasons for utilizing eMaintenance and that most solutions depend on some form of an ICT capability to fully encompass and realize all potential eMaintenance benefits, there are also an equally high probably of potential drawbacks and risks. There are no perfect solutions but rather there are always risks, especially concerning electronic network solutions.

To better understand some of the disadvantages with the use of ICT in eMaintenance, and their associated risks, this paper will focus on the potential integration of safety and security and its feasibility, i.e. is it practical, or even useful to pursue the idea of integrating safety and security for the support and protection of eMaintenance solutions. This paper will also address some previous research and noteworthy safety and security projects. Whatever the outcome from the feasibility discussion, this paper will substantiate the outcome by comparing and contrasting safety and security at a conceptual or analytical level focused on the product or system design requirements.

2. NETWORK CONNECTIVITY

The use of ICT in eMaintenance does provide a number of advantages, such as an improved overview of a system, faster responses, unified interfaces, a reduced risk of committing errors by a technician, and improved performance. Essentially, ICT is integral to the functionality and operations of eMaintenance, as shown by the eMaintenance definitions in Table 1. Based on this and the fact that ICT represents the combination of Information and Communication Technology, it is then possible to extrapolate that network connectivity is a necessary ingredient within ICT to perform its mission. Basically, if there is no network connectivity, then there is no viable eMaintenance capability. It is that simple.

So what are some examples of network connectivity within eMaintenance? To adequately address this question, it must first be understood that from the domain of network models there are two basic types of layered models; the Open Systems Interconnection (OSI) reference model and the Transmission Control Protocol/Internet Protocol (TCP/IP) model [10], [45]. Without diving too deep into these models, the lowest common

layer in these models is the physical layer, which is where this paper will mainly focus.

From a network perspective, the physical layer mainly addresses a network's topology and its associated technologies. It is at this level that network connectivity can be categorized as those technologies that are based on the use of cabling and those that are based on wireless transmission [10], [45].

There are other common layers that could also be discussed, such as the application layers, but for the purposes of this paper there is no need to do so. There exists sufficient material and need within the physical layer to adequately ascribe and justify either for or against the integration of safety and security, as it pertains to eMaintenance.

In comparing wire and wireless technologies, there are advantages and disadvantages to both. From a disadvantage stand point, the more significant issues are:

- Wireless is slower than wire;
- Wire provides less mobility than wireless;
- Wire in most cases is more expensive than wireless;
- Wireless based solutions are usually less secure than wired, although wire is still susceptible to wiretapping, which is also difficult to detect [45].

Some examples of wireless technologies that are used today in eMaintenance are ZigBee, Bluetooth, RFID, WiFi, etc. [1], [18], [25], [31], [32]. The various advantages and disadvantages of these specific technologies are nonetheless outside the scope of this paper.

Both wire and wireless network solutions provide for an increased risk of some form of interference, especially wireless solutions. The main risk that is being stressed here in this paper is directly related to the security domain of ICT. This in turn can have an impact on the safety domain. This linked effect will be discussed further in Section 5 of this paper.

3. SAFETY & SECURITY CONCEPTS

The fields of safety and security have many differences, but also many similarities. What often occurs is that the differences between safety and security become more central when discussing their use relative to a particular product or system, than their similarities. In fact, both the similarities and differences between safety and security are important, which will be discussed further in Section 5 of this paper.

For now though, there are several important similarities. First is that, safety and security basically have the same overall objective, i.e. to provide protection that leads to a condition of being free from danger or a threat [2], [3]. Another way of stating this is that there is a commonality between safety and security in terms of preventing system loss [27]. Sometimes this commonality is not intuitively evident on the surface but it becomes more apparent when following down the cause-effect chain [27]. This similarity though, implies guarding against events that are essentially negative, i.e. attacks and accidents [36]. As a consequence of this, safety and security "are often perceived as holding back productivity or more generally as obstacles to the functional requirements and aims of the systems or organizations analyzed" [36]. This perception has earned safety and security the unfavorable title as "eternal killjoys" [36].

Another similarity worth noting is that both safety and security analysis work from a base concept of risk assessment and management [14], [17], [27], [36]. This in itself has provided a basis for past and present research for the continuing endeavor of integrating safety and security as they pertain to and are

associated with various domains. Some of the more noteworthy projects and programs of the past are discussed in the next section of this paper.

Similarities are vital to understanding how the concepts of safety and security could potentially work together, but what are more important are the differences. In the past though, most of these differences have actually been portrayed as detractors and limitations to the integration of safety and security. For instance, it has been well established that safety and security are based on separate theoretical and methodological frameworks [36]. This has largely been depicted as a drawback or hindrance to integrating safety and security [36]. When in reality it is these differences that could potentially be the strength of an integrated concept of safety and security. To understand this and provide a more detailed line of reasoning, the following sub-sections focus primarily on the differences between safety and security.

3.1 Safety

There are many types of safety used today which give all the more credence to the fact that there really is no absolute definition of safety [36]. This is understandable realizing that the concept of safety has been around for many more years than security engineering.

In this light, one of the earlier documents to address safety from a management perspective was the US Military Standard 882 (MIL-STD-882) of July 1969 [33]. This document formalized the concept of system safety for the procurement of products and systems within the US Military. As a result of this document, it is mainly from the military domain that system safety initially grew and matured to what it is today. From the current revision of this document, MIL-STD-882E, safety has been defined as "freedom from conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment" [33]. System safety is defined as the "application of engineering and management principles, criteria, and techniques to achieve acceptable risk within the constraints of operational effectiveness and suitability, time, and cost throughout all phases of the system life-cycle" [33]. Since this paper is focused on eMaintenance from an engineering and management perspective, the use of the term "safety" from here on will mean the MIL-STD-882E system safety definition.

With this being formulated and as previously stated in Section 3 of this paper, safety utilizes different theoretical and methodological frameworks for its analysis. More specifically, safety predominately uses quantitative methods that are more based on probabilistic approaches [36]. These quantitative methods can be characterized as being either (1) practical tests and through a series of measurements, or (2) by using abstract models [36].

It was mentioned earlier in this section that both safety and security analysis work from a base concept of risk assessment and management. This though is where most of the similarities end.

Risk management within safety is mainly focused on hazard identification, elimination, and control through the use of mitigation techniques [15]. As such, hazards are defined in the MIL-STD-882E as:

A real or potential condition that could lead to an unplanned event or series of events (i.e. mishap) resulting in death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment.

Essentially, a hazard is comprised of three elements: (1) hazardous element, (2) initiating mechanism, and (3) target and threat [15]. There is a potential for confusion here in that the “target and threat” as it is used with respect to a hazard, refers to a person or object that is susceptible to injury and/or damage, and describes the severity of the mishap event [15]. There is no association with the security threat terminology. The focus of this information is to show that this type of risk analysis is in fact different from than that of security.

Another distinction with the concept of safety is that of intent, i.e. there is no malicious intent involved with a safety analysis [36]. Only accidents are addressed. Thus, from a safety perspective, an accident is the result of non-malicious failure. This then provides a clearer correlation for the feasibility of utilizing quantitative methods of analysis. Further comparison and elaboration on this topic is discussed in Section 4.1 on the SEMA referential framework [37]. As a result of this is the concept of a fault which has been defined by Ericson as:

“...the occurrence or existence of an undesired state for a component, subsystem, or system. In a fault, the component operates correctly, except at the wrong time, because it was commanded to do so (due to the system design)” [16].

The concept of safety today is widely accepted as a necessity to the successful and error free operations of almost any product or system design. To accomplish this, safety processes, procedures, and considerations are integrated into the product or system design as early as possible and at all levels. In addition to this, safety also has to consider outside influences from its environment but only non-malicious considerations. For the purposes of this paper, the integration of all safety processes, procedures, considerations, and risk mitigations within a particular product or system are referred to as “safety measures,” as depicted in Figure 2, below.



Figure 2. Conceptual relationship between safety measures and its environment

One of the challenges with two-dimensional drawings, like the top down perspective in Figure 2, is that they often loose or over shadow other potentially significant relational aspects. In this case, there is no simple way to depict the relationship between a product’s/system’s design and its associated safety measures. To solve this challenge, a side perspective is provided in Figure 2.

In this way, Figure 2 provides a layered perspective, whereby it is now possible to visualize the relationship between the product or system design and its associated safety measures, called a safety gap. Ideally, if all required safety measures are correctly implemented and integrated in the associated product or system design, there would be no space in the safety gap, but rather a completely integrated product or system, with respect to safety measures and product or system design.

The use of this type of an extrapolation is simple, at this point, but this conceptualization will prove to be more of value later in this paper. This will be accomplished by building upon and developing a more detailed and comprehensive justification for the potential need to integrate safety and security as it applies to the eMaintenance domain.

There are other differential distinctions of safety relative to security, but the aforementioned are, for the most part, the more significant to this paper.

3.2 Security

Security, as it pertains to the field of engineering, is a relatively new domain compared to safety. Security engineering is the result of the drive towards interconnectivity and interoperability of today’s networks, computers, applications, and even enterprises [21]. The field of security engineering continues to grow and mature rapidly, although considered by some, like Bruce Schneier, to be in a game of perpetual catch-up [41].

As a discipline, security engineering interfaces with many other disciplines such as system engineering, software engineering, human factors engineering, and communications engineering [21]. Within security engineering itself, there are many specialties or security sub-disciplines such as operations security, information security, network security, physical security, and communication security [21]. The collective use of these security sub-disciplines as a bridge between security engineering and systems engineering is called Systems Security Engineering [21]. Tinney and Candell [44] state that the use of Systems Security Engineering is advocated because it provides for and assists with organizing and understanding a product’s or system’s requirements and processes as they pertain to: the entire life cycle, the entire organization, concurrent interactions with other disciplines, and all security sub-disciplines.

As a field of research though, security engineering is vast with many specialties, but also many commonalities. It is these commonalities that this paper will now focus; thereby inferring that all further usage of the “security” terminology will imply security engineering.

From a general point of view, “security is the process of maintaining an acceptable level of perceived risk” [7]. There are other definitions that may apply, but this definition is quite appropriate in that it is succinct. As a process, this definition implies that security is not an obtainable objective or achievable [7]. Basically, the state of being secure is only as good as the last verification.

The thought of having an objective that is not achievable is sometimes difficult for many in management to feel comfortable with. So why is complete security not achievable?

To adequately address this question, it must first be understood that security risk is comprised of three main components; threats, vulnerabilities, and impacts. From ISO/IEC 21827 [21] these terms are defined as:

Threat - capabilities, intentions and attack methods of adversaries, or any circumstance or event, whether originating externally or internally, that has the potential to cause harm to information or to a program or system, or to cause these to harm others.

Vulnerability - includes a weakness of an asset or group of assets which can be exploited by a threat.

Impact - the result of an uncertain and unwanted incident from a threat exploiting a vulnerability.

With this being stated, the objective of risk management within security is to reduce risk by identifying problems that have not yet occurred [21]. This is more difficult than it sounds. The main challenge here is how to handle and control the uncertainty with an every changing, thus dynamic threat. Schneier emphasizes that the known threats of today will not be the same threats of tomorrow [41]. This is the primary reason why no product or system should be considered completely secure at any time.

Understanding the security risk process is essential to understanding how security countermeasures are used and developed, i.e. a countermeasure is an action, device, procedure, technique, or other measure that reduces the vulnerability of a product or system towards a threat [19], [43]. This is a generic definition that also implies that a specific countermeasure can also be designed to provide protection against additional unspecified or unknown threats [43].

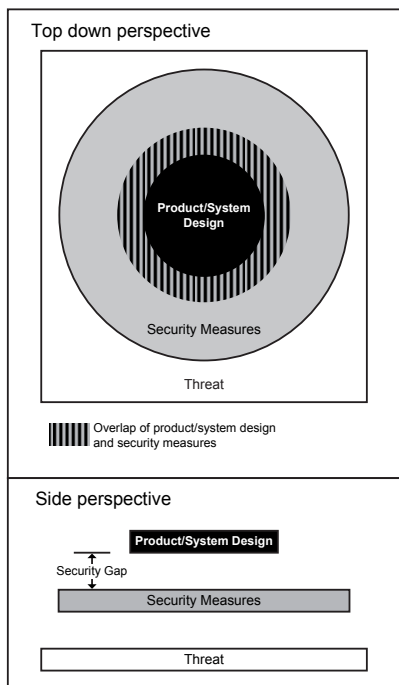


Figure 3. Conceptual relationship between security measures, the product/system design, and the threat

For the purpose of this paper, the collective integration of all security processes, procedures, countermeasures, and risk mitigations relative to a particular product or system, will now be referred to as “security measures,” as depicted in Figure 3, below. In this depiction, the security measures layer provides countermeasures to thwart attacks from the threat. This in many ways is one of the primary goals for the security measures, i.e. to stop the external threat.

Also unique to Figure 3, is the overlap between security measures and the product/system design. As delineated by Cockram and Lautieri, similarities actually produce overlaps [12]. In this and the next conceptualization, the overlap represents:

- The span or width of the overlap directly corresponds to the amount of integration between the applicable security measures and product/system design;

- All processes, procedures, considerations, and risk mitigations designed to thwart the internal security threat.

Another consideration for an overlap is to represent the fact that not all of a product’s or system’s design pertains to or is concerned with the provisioning of security measures. In other words, not all of the product/system design is required or needed in coordinating the support for the security measures. For the most part, the only time there is an inward focus within a product or system is to address the internal security threat, which is a difficult task in itself [12].

With respect to the security gap shown in Figure 3, this represents the relationship between the product/system design and the associated security measures. Again, ideally, if all required product/system design requirements are correctly implemented and integrated with the security measures, there would be no gap between them.

In the discussion on safety in Section 3.1, there are other differential distinctions of security relative to safety, but the aforementioned are, for the most part, the more relevant to this paper.

Before discussing the potential necessity of integrating the safety and security domains, from the perspective of the authors, we first briefly present four past projects and their different goals and ambitions. These projects are discussed with respect to the safety-security integration domain.

4. PAST NOTEWORTHY PROJECTS

There have been several projects within the last 20 years that have focused on either categorizing or attempting to integrate some aspects of safety and security. The following is a concise review of some of the more prominent and noteworthy projects that have been chosen due to their pertinence to the scope of this paper. If a project is not mentioned below, it by no means infers any judgment against or for that project.

4.1 SEMA

The first project to be discussed is the SEMA Referential Framework. SEMA stands for System vs. Environment and Malicious vs. Accidental [37] and was first described in the International Journal of Critical Infrastructure Protection in 2010. It was developed because of the varied meanings of the terms safety and security. Basically, depending on how these terms are used and in what contextual domain, their use can be problematic, if nothing else because of potential ambiguities [37]. The SEMA Referential Framework makes some of the latent differences in using the terms safety and security more explicit to reveal potential inconsistencies and overlaps [36], [37].

The framework consists primarily of two distinctions, the (1) System vs. Environment (and vice versa) and the (2) Malicious vs. Accidental [37]. These distinctions were derived mainly from analyzing many sets of definitions as they pertain to approximately 12 different sectors [37]. A visual representation of SEMA is seen in Figure 4.

In the interest of brevity, some of the more detailed inner structure of the SEMA Referential Framework will be bypassed. What is pertinent to understand is that the SEMA framework structure has been augmented, for the sake of completeness, to include a system-to-system dimension [37].

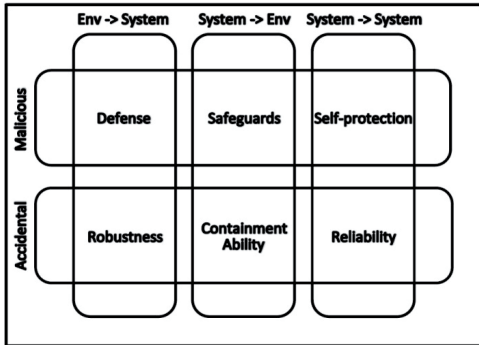


Figure 4. SEMA referential framework from Piètre-Cambacédès and Chaudet [37]

As depicted in Figure 4, the SEMA framework is divided into six distinct sub-notions: defense, safeguards, self-protection, robustness, containment ability, and reliability [37]. The authors “argue that the six sub-notions are semantically less ambiguous than the generic terms security and safety, and that they consistently cover their conceptual domains” [37]. It is however, not the intention of SEMA to replace the safety and security terms, but rather to help establish a common understanding between technical communities [37]. SEMA is also “useful during the early stages of system design and when defining the scope of a risk assessment” [37].

The use of the sub-notions does require that system boundaries be clearly identified and explicitly stated [37]. There also maybe a misperception that the sub-notions are mutually exclusive. In reality, an event or technical measure can actually span several sub-notions.

Another limitation of SEMA is that it cannot solve intrinsic problems from imprecise or inherently overlapping definitions of safety and security in some domains [37]. SEMA can, although, help with identifying inconsistencies and overlaps with the use of safety and security conceptualizations, thereby help save time and resources [37].

The SEMA referential framework is useful in assisting in the design process but it must be emphasized that it is not a design process in itself. If attempting to integrate safety and security within the context of a product or system design, the SEMA referential framework can be of value if utilized appropriately.

4.2 SafSec

SafSec stands for Safety and Security. It was a project sponsored by the British Defence Procurement Agency from approximately 2004 to 2006 in an effort to manage and reduce certification and approval costs [12], [26], [27]. It aimed to support safety certification and security accreditation of

complex computer based systems, in particular those deployed as Integrated Modular Avionics systems [27].

SafSec focused essentially on two major issues [27]:

- Identifying and exploiting commonalities between various certification processes, and;
- Providing a framework for the certification of modular systems (systems composed of standard components which are re-used in different configurations and applications).

The project found commonalities and overlaps between safety and security [27]. As such, there were commonalities on the surface from a general perspective, but when pursuing a more detailed analysis the results often showed a larger variation of differences such as in the analysis methods and mitigation processes.

The results of the SafSec project are as follows [27]:

- A unified approach to risk assessment for safety and security for certification processes;
- A risk directed design process, which includes risk mitigation decisions in the design process and produces substantiated arguments to support them;
- A process supporting certification of modules within a modular architecture.

These results were and are of value but primarily as they are applied to certification processes. There were some developments with a Risk-Directed Design Framework but the methodology guidance material did not provide sufficient details to make a valued judgment to the frameworks usability [13]. Even though there were some product or system level design considerations within SafSec, there still was no common design process that incorporates both safety and security analyses, either sequentially or concurrently.

4.3 SEISES

SEISES is a French collaborative project that is an abbreviation for Systemes Embarqués Informatisés, Sûrs Et Sécurisés, which essentially means Safe and Secure IT Embedded Systems [9]. The objectives of the SEISES project were to:

- Elaborate on a harmonized process to support the development and validation of systems with both safety and security requirements in a cost effective manner, and to;
- Support the elaboration of the proper justification evidence as required by the applicable safety and security standards and authorities [9].

One of the results of the SEISES project was the development of a safety and security assurance framework [8]. The basic principles used to design this framework were:

- Processes should deal with “security for safety”;
- Processes should focus on development assurance;
- Processes should be described by a set of assurance objectives;
- Assurance objectives should be implemented using assurance activities described in existing standards, and;
- Convergence between safety and security assurance activities should be described [8].

Using these design principles, the SEISES framework provided a collection of safety and security assurance objectives that are consistent with current practices in safety critical system development [8]. Examples of safety critical systems are products or systems whereby failure is to be avoided if at all possible such as those products or systems associated with an

airplane's flight control capability [4], [39]. From 2008 to 2011, the SEISES project partners used the resultant framework in the development of secure and safe embedded aerospace systems [8].

Here again is some noteworthy research in the use of safety and security principles combined to produce viable processes and procedures as they apply to safety and security assurance. This still though do not address or have a direct bearing on any capability to integrate safety and security at the product or system design level.

4.4 SQUALE

The SQUALE (Security, Safety and Quality Evaluation for Dependable Systems) project ran from 1996 to 1998 [42]. The objective of the project was to analyze the existing standards and practices in the safety area, as well as security, and define a combined harmonized approach to gain confidence in the correctness and effectiveness of systems with safety and security requirements.

One problem that arose within this framework was that it tried to cover both safety and security requirements and functions that originated from the two domains such that they may overlap, or even be contradictory. Hence, a simple combination of practices used in the safety domain and used in the security domain would not necessarily guarantee a design that would satisfy both requirements. Therefore, a harmonized approach was needed to simultaneously address both safety and security requirements [42].

The project defined a "dependability criteria" that described this harmonized approach and applied these criteria to a demonstrator system [42]. The demonstrator system, in this case, was the Automated Train Operation System (ATOS), which was developed by Matra Transport to equip the METEOR (MÉTro Est-Ouest Rapide) underground line of the Paris metro. The existing system had very high dependability requirements and was therefore the ideal candidate to test the criteria developed within the project.

The new criteria were based on the many existing standards in the safety and security arenas and did take into account the entire system life cycle. A mapping to existing standards (which covered only specific dependability aspects) was provided as part of the project. Moreover, the new criteria was general enough to allow systems to be evaluated according to domain specific standards (e.g. IT security and International Electrotechnical Commission (IEC) 1508) and other application specific sector standards (e.g. Radio Technical Commission for Aeronautics DO-178B / European Organisation for Civil Aviation Equipment ED-80 *Software Considerations in Airborne Systems and Equipment Certification*, and European Standard EN-50126 *Railway applications – The specification and demonstration of Reliability, Availability, Maintainability and Safety*) without requiring too much extra effort and cost.

Herewith, the SQUALE evaluation framework and criteria incorporated some basic concepts from various existing safety standards. More specifically:

- The roles of the different parties involved in the evaluation process;
- The notion of target of evaluation;
- A process oriented evaluation framework;
- Different levels of confidence to grade the importance of the dependability attributes;
- Different levels of rigor, detail and independence, in respect to each individual dependability attribute.

One important observation in the SQUALE project was that in many situations it is not possible to determine if a product or system failure is caused deliberately, accidentally or by an equipment fault [42]. The consequences are nonetheless identical. Furthermore, the risk mitigations proposed by both the safety and security sectors are very similar. This observation and the potential confusion that can result from an overlap of safety and security requirements further reinforces the supposition for the need of a viable integrated safety and security product or system design methodology.

4.5 Summary of Past Projects

The term integration can have various meanings. In this paper integration means either to unify safety and security concepts or to harmonize safety and security concepts. Examples of four projects that have been focusing on either one of the two concepts have been presented above. The following summarizes the four projects in relation to the two integration concepts.

4.5.1 Unification

One of the SEMA Referential framework's [37] main contributions was that of attempting to unify the domains of safety and security. As it was stated earlier, the framework could be of assistance in the context of a product or system design, if utilized appropriately. The project was also ambitious when embarking on a system-to-system dimension. The downside of the framework is its difficulty to handle the potential overlaps between safety and security concepts.

SafSec [27] was another project that utilized a unification approach. This project had a limited objective in that it mainly focused on the requirements associated with the certification process (safety) and the accreditation process (security). Thus, the ambition of the project was to some extent more confined than that of the SEMA project within the domain of safety and security integration.

In many senses, the SEISES project [9] has some similarities to that of the SafSec project as its emphasis was more on assessment. However, the project had another focus which was on the assurance of development processes. But again, as was the case in the two aforementioned projects, SEISES did not address the possibility of integrating safety and security at the product or system design level.

4.5.2 Harmonization

The last project that was briefly presented was SQUALE [42]. This project was focused on attempting to harmonize, not unify, the safety and security domains by simultaneously addressing safety and security requirements. The project pointed at the risk of a potential confusion that could come as a result of an overlap in the safety and security requirements. This risk of confusion is where the authors of this paper disagree. We fully understand the need to be observant of these overlaps and that they in some respects can be detrimental to the functionality of a system. As such, it will be discussed in the next section that the inherent differences between the two domains could potentially be a strength and justification for integrating, and more specifically harmonizing, the two domains.

5. INTEGRATION OF SAFETY & SECURITY

As previously mentioned, there are two primary methods of integration in this paper, either to unify concepts or to harmonize concepts. The unification approach to integrating safety and security has been investigated by McDermid in 1994 [30] and by Sanders and Meyer in 1991 [40], whereby safety and security were unified under the heading of dependability and included with other attributes such as reliability and

availability [14]. More recently this methodology was utilized and developed further by Avižienis et al. in 2001 [5] and again in another paper in 2004 [6]. Also the SafSec project stresses the need to “establish sufficient dependability” [27]. Having stated this, it is believed that the “dependability path” is not the most viable path to obtain tangible results when it comes to a integrating the two domains.

The other integration approach is harmonization. This is where the differences and/or similarities between safety and security are examined, such that one or several differences could be of some value if utilized by the other concept. More recently, Piètre-Cambacédès and Bouissou presented some examples of work related to harmonization approaches in their paper *Cross-fertilization between safety and security engineering* [36]. For the remainder of this paper, the term integration will refer to the harmonization approach.

The next question to address, at this point, is whether or not it is practical, or even useful to pursue the idea of integrating safety and security for the support and protection of eMaintenance solutions? To effectively address this question, there needs to be a two part answer. The first part of the answer will address an integrated safety only solution for the protection of an eMaintenance solution; and the second part will address the situation of an integrated security only solution.

5.1 Safety Only Solution

As discussed in Section 3.1, of this paper, safety processes, procedures, and considerations do not cover instances or events that are the result of malicious intent. Piètre-Cambacédès and Bouissou mention this specifically [36]. In many respects a safety only solution for the protection of any product or system that uses any form of network connectivity, is only asking for problems. From an IT design perspective, Anderson in his book *Security engineering: a guide to building dependable distributed systems* [4] states that comparatively, safety is like programming Murphy’s computer, where security is akin to programming Satan’s computer.

This in effect though is not bad. This difference of intent should not be viewed as a disadvantage, but rather as an incentive or as an asset. Safety processes and procedures work because they have specialized techniques that “have evolved with the aim of producing a thorough and complete analysis” [14]. To dilute these well-established processes and procedures would be a step in the wrong direction.

5.2 Security Only Solution

The security only solution, as described in Section 3.2 of this paper, is primarily focused on risks that originate from or exacerbated by way of malicious intent [36]. As shown in Figure 3, not all of the security measures pertain to the entire product/system design. In effect, not all of the product/system design needs to be involved with protection against the threat. From a control only point of view, this could be perceived as a limitation or disadvantage of the security measures layer.

A more realistic security point of view is that if less is good enough, more is usually not better. This is especially true when dealing with lines of software code. Fewer lines of code mean fewer risks.

This still does not solve how to assure that the internal functionality of a particular product or system will perform as designed. A security only solution just does not provide the

needed processes, procedures, and considerations to accomplish this.

5.3 The Integrated Solution

There have been several significant attempts to integrate some aspect of safety and security using either a unification approach, harmonization approach, or a combination of both. Some of these attempts were discussed in Section 4 of this paper. In the end though, there have been no conclusive attempts to integrate safety and security that has been universally accepted throughout today’s industrial base. One of the primary reasons for this is the uniqueness and complexity of each and every product or system and their associated requirements [4], [16]. No one solution will satisfy all requirements from either a safety or security perspective.

As discussed in Sections 5.1 and 5.2, the safety only and security only solutions, respectively, do not provide the needed comprehensive protective solution required to support a network connected eMaintenance environment of today. There has to be a better methodology.

When reading recent publications it becomes evident that most modern complex systems currently developed only consider one of the two concepts, either safety or security during the design phase. Limiting the design by contemplating only one of the two concepts significantly increases the risk of incurring one or more serious problems later in the product or system life cycle.

Another aspect to consider is which direction technology is heading towards over the next decade or more. Using the last decade as a gauge, it has been quite evident that one of the main technological driving forces has been an ever increasing requirement for more network connectivity [38]. Thus the need for product or system solutions will only increase.

Conceptually though, how would an integrated safety and security solution look? One of the ways to visualize this is shown in Figure 5, below. In this conceptualization and as previously stated in Section 3.1, the safety measures and product/system design ideally are a single layer, but for this conceptualization, they are also separated. Using this idea, the overlap is between the safety measures and security measures then represents the provisioning of a mutually supporting integrated protective capability against both the internal and external threat.

As such, it could be reasonably inferred that not all of the environment that a particular product or system comes in contact with is actually a threat. This is true, but in the case of product or system design, it must be stressed that the need should be to plan worst case, i.e. consider everything in the environment, first and foremost as a threat. This should be applied not only to the external environment, but also the internal environment.

The safety and security gap depicted in Figure 5 now represents those issues not sufficiently rectified and collaboratively harmonized. As more and more integration occurs, there will be less and less gap between the safety and security measures.

This conceptualization could be a viable planning tool when designing eMaintenance solutions, not only for designing network connectivity functionalities and capabilities, but also as a tool to address all potential eMaintenance protective requirements. In other words, this conceptualization represents an integrated holistic protective planning approach to eMaintenance.

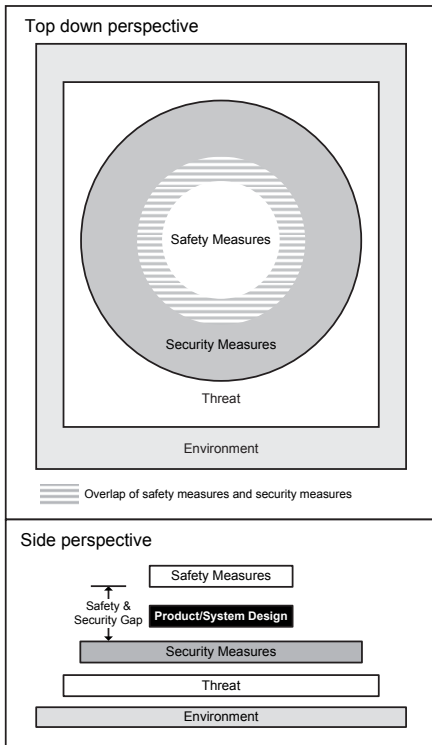


Figure 5. Conceptualization of the relativity of safety and security combined

6. FUTURE WORK

The conceptualizations in this paper have not been validated, as of the time of this paper. Although, there have been many discussions over the last couple of years between various safety and security practitioners that have significantly assisted in conceiving and developing these conceptualizations.

Even once these conceptualizations have been validated, there is much that can and should be accomplished to systematize the integration of safety and security. In such a case, a framework to help organize this type of effort would be of value.

Be aware though, much of the terminology used in this paper has its origins from the aviation domain and may need some slight modifications to be applied to other subject domains. One of the reasons for using the aviation domain as a starting point is due to the fact that the aviation domain has heavily invested large sums of resources into their safety measures and processes. Although this can be considered a limiting factor, it does assure that the safety processes and procedures currently being used have been heavily scrutinized. With this, security as within the aviation domain as a whole has been relegated to a supporting role. This though is not always the case for other potential subject domains.

7. CONCLUSIONS

The focus of this paper has been to increase the awareness relative to the potential disadvantages and drawbacks with the use of ICT. These drawbacks are often forgotten or sometimes even overlooked within systems engineering, let alone in

eMaintenance. The benefits that ICT provides in relation to the maintenance of complex systems are widely recognized and utilized in many projects, as exemplified by the maintenance activities related to a fighter aircraft [11], to the lowest level dealing with simple artefacts such as sensors or tags [1], [18], [31], [32]. One limitation with these examples, and many other examples similar to these, is a lack of planning with respect to the drawbacks with utilizing ICT, and more specifically those related to security. The importance of safety and the role it plays in maintenance is comprehended by most of the maintenance community but security to a much lesser extent. Failing to realize the importance of the security domain and its intrinsic intertwining with the safety domain could have a profound and negative impact on the usefulness and positive effects that otherwise come from ICT in eMaintenance.

This paper began with a question of whether it would be potentially practical or even useful to pursue the idea of integrating safety and security for the support and protection of eMaintenance solutions. To gain a better understanding of some of the precepts associated with eMaintenance, a series of eMaintenance definitions were investigated. In reviewing these definitions, it was found that one of the primary requirements to provide an eMaintenance capability was the use of ICT.

By decomposing the ICT capability, it was shown that the network connectivity functionality was a significant source of potential vulnerabilities. This was the case regardless of the technology that was utilized, i.e. wire or wireless usage.

It was then conceptualized and investigated what and how a safety only and a security only solution might be of use relative to eMaintenance network connectivity solutions. It was later concluded that both solutions were inadequate to provide any type of comprehensive protective capability by their own. This idea, in many ways, has been concluded by many other projects. What is different though is that the differences between safety and security should actually be considered strengths and utilized as such. As pointed out by Eames and Moffet, is that by attempting to unify safety and security could very well dilute specialized techniques that have evolved to provide a thorough and complete analysis [14].

As a result of these and other considerations, like complexity, it was concluded that there is more than a need for an integrated safety and security capability within eMaintenance. In fact, an integrated safety and security capability is actually an imperative. So the answer to the first question, *whether or not it is practical, or even useful to pursue the idea of integrating safety and security for the support and protection of eMaintenance solutions*, is a resounding yes as demonstrated in this paper. The answer to the supplementary question *are there any potential synergies that could be capitalized on to provide a more thorough and comprehensive protective capability for most eMaintenance solutions against security threats* an example was given, stating that the overlap between the safety measures and security measures represents the provisioning of a mutually supporting integrated protective capability against both internal and external security threats.

ACKNOWLEDGMENTS

This work has been partially supported by the Swedish National Aeronautics Research Programme, NFFP5 Call 2.

8. REFERENCES

- [1] Adgar, A., A. Yau, and D. Addison (2010). "Smart tags" (book chapter 8), *eMaintenance*, Springer, book edited by Holmberg, K., A. Adgar, A. Arnaiz, E. Jantunen, J. Mascolo, and S. Mekid, pp. 197-225.

- [2] Albrechtsen, E. (2002). "A generic comparison of industrial safety and information security", *Term paper in the PhD course "Risk and Vulnerability"*, NTNU. December 2002. Available at: <http://www.iot.ntnu.no/users/albrecht/rapporter/generic-comparison-of-ind-saf-and-inf-sec.pdf>, accessed: 1 April 2014.
- [3] Albrechtsen, E. (2003). "Security vs safety", Norway: NTNU. August 2003. Available at: <http://www.iot.ntnu.no/users/albrecht/rapporter/notat-safety-v-security.pdf>, accessed: 1 April 2014.
- [4] Andersson, R. (2008). "Security engineering: a guide to building dependable distributed systems", 2nd edition, *John Wiley & Sons*, U.S.A.
- [5] Avižienis, A., J.C Laprie, and B. Randell (2001). "Fundamental concepts of dependability", *Technical report 739*, Department of Computing Science, University of Newcastle upon Tyne Computing Science.
- [6] Avižienis, A., J.C. Laprie, B. Randell, and C. Landwehr (2004). "Basic concepts and taxonomy of dependable and secure computing", *Dependable and Secure Computing*, 1(1), January-March 2004.
- [7] Bejtlich, R. (2005). "The Tao of network security monitoring: beyond intrusion detection", *Addison Wesley*.
- [8] Bieber, P., J.-P. Blanquart, G. Descargues, M. Dulucq, Y. Fourastier, E. Hazane, M. Julien, L. Léonardon, and G. Sarouille (2012). "Security and safety assurance for aerospace embedded systems", *The 6th European Congress on Embedded Real-Time Software and Systems*, pp. 1-10, 1-3 February 2012, Toulouse, France.
- [9] Blanquart, J.-P., P. Bieber, and G. Descargues (2012). "Similarities and dissimilarities between safety levels and security levels", *The 6th European Congress on Embedded Real-Time Software and Systems*, 1-3 February 2012, Toulouse, France.
- [10] Bragg, R., M. Rhodes-Ousley, and K. Strassberg (2004). "Network security: the complete reference", *McGraw-Hill*, New York, U.S.A.
- [11] Candell, O., R. Karim, and P. Söderholm (2009). "eMaintenance – Information logistics for maintenance support", *Journal Robotics and Computer-Integrated Manufacturing*, 25, pp. 937-944.
- [12] Cockram, T.J. and S.R. Lautieri, (2007). "Combining security and safety principles in practice", *The 2nd Institution of Engineering and Technology Conference on System Safety 2007*, pp. 159-164, 22-24 October 2007, Toulouse, France.
- [13] Dobbing, B. and S. Lautieri (2006). "SafSec methodology: guidance material", *Praxis High Integrity Systems 2005*. Available at: http://intelligent-systems.altran.com/fileadmin/medias/0.commons/documents/Technology_documents/SafSec_Methodology_Guidance_Material_pdf.pdf, accessed: 1 April 2014.
- [14] Eames, D.P. and J. Moffett (1999). "The integration of safety and security requirements", *SAFECOMP'99, LNCS 1698*, Springer-Verlag, pp. 468-480.
- [15] Ericson, C. (2005). "Hazard analysis techniques for system safety", *John Wiley & Sons*, Hoboken, New Jersey, U.S.A.
- [16] Ericson, C. (2011). "Concise encyclopedia of system safety: definition of terms and concepts", *John Wiley & Sons*, Hoboken, New Jersey, U.S.A.
- [17] Firesmith, D.G. (2005). "A taxonomy of security related requirements", *Software Engineering Institute*, White Paper, May 2005.
- [18] Halme, J., E. Gorritxategi, and J. Bellew (2010). "Lubricating oil sensors" (book chapter 7), *eMaintenance*, Springer, book edited by Holmberg, K., A. Adgar, A. Arnaiz, E. Jantunen, J. Mascolo, and S. Mekid, pp. 173-195.
- [19] Herrmann, D.S. (2007). "Complete guide to security and privacy metrics". *Auerbach Publications*, U.S.A.
- [20] Ismail, S., M. Sarangi, A. Parida, and U. Kumar (2010). "Cause failure analysis of coal mill vertical shaft", *The 1st International workshop and congress on eMaintenance*, pp. 176-181, 22-24 June, 2010, Luleå, Sweden.
- [21] ISO/IEC 21827 (2008).
- [22] Kajko-Mattsson, M., R. Karim, and A. Mirjamdotter (2010). "Fundamentals of the eMaintenance concept", *The 1st International workshop and congress on eMaintenance*, pp. 147-154, 22-24 June, 2010, Luleå, Sweden.
- [23] Karim, R. (2008). "A service-oriented approach to eMaintenance of complex technical systems", *Doctoral thesis 2008:58*, Luleå University of Technology, Sweden.
- [24] Karim, R. and P. Söderholm (2009). "Application of information and communication technology for maintenance support information services", *Journal of Quality in Maintenance Engineering*, 15(1), pp. 78-91.
- [25] Krommenacker, N., V. Lecuire, N. Salles, S. Katsikas, C. Giordamli, and D. Emmanouilidis (2010). "Wireless communication" (book chapter 10), *eMaintenance*, Springer, book edited by Holmberg, K., A. Adgar, A. Arnaiz, E. Jantunen, J. Mascolo, and S. Mekid, pp. 247-271.
- [26] Lautieri, S. (2006). "De-risking safety", *IET Computing and Control Engineering*, June/Jul 2006 pp.38-41.
- [27] Lautieri, S., D., Cooper, and D. Jackson (2005). "SafSec: commonalities between safety and security assurance", *Proceedings of the 13th Safety Critical Systems Symposium*, Southampton, February 2005, Springer-Verlag London Ltd, Praxis Critical Systems.
- [28] Lee, J., J. Ni, D. Djurdjanovic, H. Qiu, and H. Liao (2006). "Intelligent prognostics tools and e-maintenance", *Computers in Industry*, 57, Elsevier, pp. 476-489.
- [29] Lin, J.-S., Y.-Y. Chang, C.-Z. Liu, and K.-W. Pan (2011). "Wireless sensor networks and their applications to the healthcare and precision agriculture" (book chapter 15), *Wireless Sensor Networks*, InTech, book edited by Tarannum, S.
- [30] McDermid, J.A. (1994). "On dependability, its measurement and its management", *High Integrity Systems*, 1(1), pp. 17-26.
- [31] Mekid, S., A. Starr, and R. Pietruszkiewicz (2010). "Intelligent wireless sensors", (book chapter 5), *eMaintenance*, Springer, book edited by Holmberg, K., A. Adgar, A. Arnaiz, E. Jantunen, J. Mascolo, and S. Mekid, pp. 83-123.
- [32] Mekid, S. and Z. Zhu (2010). "MEMS sensors" (book chapter 6), *eMaintenance*, Springer, book edited by Holmberg, K., A. Adgar, A. Arnaiz, E. Jantunen, J. Mascolo, and S. Mekid, pp. 125-171.
- [33] MIL-STD-882E (2012). *Department of Defense Standard Practice System Safety*. US Department of Defense.

- [34] Muller, A., A.C. Marquez, and B. Iung (2008). "On the concept of eMaintenance: review and current research", *Journal of Reliability Engineering and System Safety*, 93(8), Elsevier Science Publishers B.V., pp. 1165-1187.
- [35] OECD (2010). "Smart sensor networks for green growth", *OECD Information Technology Outlook 2010*, OECD Publishing, pp. 227-256.
- [36] Piètre-Cambacédès, L. and M. Bouissou (2013). "Cross-fertilization between safety and security engineering", *Reliability Engineering and System Safety*, 110(2013), Elsevier, pp. 110-126.
- [37] Piètre-Cambacédès, L. and C. Chaudet (2010). "The SEMA referential framework: avoiding ambiguities in the terms "security" and "safety""", *International Journal of Critical Infrastructure Protection*, July 2010, 3(2), pp. 55-66.
- [38] Sadvandi, S., N. Chapon, and L. Piètre-Cambacédès, (2012). "Safety and security interdependencies in complex systems and SoS: challenges and perspectives", *Complex Systems Design and Management*, chapter 16, Springer, pp. 229-241.
- [39] Sampigethaya, K, R. Poovendran, and L. Bushnell (2008). "Secure operation, control, and maintenance of future E-enabled airplanes", *Proceedings of the IEEE*, December 2008, 96(12), pp. 1992-2007.
- [40] Sanders, W.E. and J.F. Meyer (1991). "A unified approach to specifying measures of performance, dependability and performability", *Dependable Computing for Critical Systems*, Springer-Verlag. Wien, Austria.
- [41] Schneier, B. (2006). "No fear", Springer.
- [42] SQUALE (1999). ACTS95/AC097, End of Project Report 26 February 1999.
- [43] Talbot, J. and M. Jakeman (2009). "Security risk management – Body of knowledge", *John Wiley & Sons*, U.S.A.
- [44] Tinney, T. and O. Candell (2012). "Applying an integrated systems security engineering approach within aviation eMaintenance", *The 2nd International workshop and congress on eMaintenance*, pp. 43-48, 12-14 December, 2012, Luleå, Sweden.
- [45] Tipton H. and K. Henry (2007). "Official (ISC) 2 Guide to the CISSP CBK", *Auerbach Publications*, U.S.A.
- [46] Tsang, A.H.C. (2002). "Strategic dimensions of maintenance management", *Journal of Quality in Maintenance Engineering*, 8(1), pp. 7-39.
- [47] Zhao, Z., F.-L. Wang, M.-X. Jia, and S. Wang (2010). "Predictive maintenance policy based on process data", *Chemometrics and Intelligent Laboratory Systems*, 103, Elsevier, pp. 137-143.

Health and Performances Machine Tool Monitoring Architecture

Agustin Prado¹, Aitor Alzaga², Egoitz Konde³, Gabriela Medina-Oliva⁴, Maxime Monnin⁵, Carl-Anders Johansson⁶, Diego Galar⁷, Dirk Euhus⁸, Mike Burrows⁹, Carlos Yurre¹⁰

¹GORATU M-H, Spain, agustin.prado@goratu.com

²IK4-TEKNIKER, Spain, aitor.alzaga@tekniker.es

³IK4-TEKNIKER, Spain, egoitz.konde@tekniker.es

⁴PREDICT, France, gabriela.medinaoliva@predict.fr

⁵PREDICT, France, maxime.monnin@predict.fr

⁶LULEA TEKNISKA UNIVERSITET, Sweden, carl-anders.johansson@ltu.se

⁷LULEA TEKNISKA UNIVERSITET, Sweden, diego.galar@ltu.se

⁸ARTIS, Germany, dirk.euhus@artis.marposs.com

⁹MONITION, England, M.Burrows@monition.com

¹⁰FAGOR AOTEK, Spain, carlos.yurre@aotek.es

ABSTRACT

In order to face the high market competitiveness, the Power-OM project (<http://power-om.eu/>) aims at implementing a proactive approach for improving the machine tool performances. For implementing a proactive approach that helps monitoring machine tool performances, this paper presents a technical architecture with two levels: the local and the remote one. In the local level, condition based maintenance strategy is implemented and real time data is used for monitoring the local health of a machine. A first originality is to use the current analysis for assessing the health status of the machine. In the remote level, offline data is stored in an eMaintenance platform, which allows providing a fleet dimension. This dimension allows to benefit of more data and information allowing to make performances comparison across the fleet and along the time.

This paper presents the advantages of the added-value architecture. On one hand, there is the possibility to track performances to detect drifts locally and real-time based on the current analysis, and on the other hand, to follow-up short and mid-term performances for deeper analysis of the fleet performances in order to bring relevant information for decision makers.

Keywords

Health assessment, energy assessment, fingerprint, maintenance, fleet-wide, process monitoring

1. INTRODUCTION

Production systems deteriorate with the usage and age. The normal strategy to keep them in good conditions is to apply preventive maintenance practices, with a supportive workforce “reactive” in the case of clearly detected malfunctions or machine breakdowns. All these have an impact on quality, cost and in general, productivity. Added to this, the uncertainty of machine

reliability at any given time, also impacts on product/production delivery times. It is known also that a worn-out or incorrectly assembled mechanism has higher energy consumption.

The use of intelligent predictive technologies and tools could contribute to improve the situation (Lung et al., 2009). The deployment of these tools and techniques can detect malfunctions or potential breakdowns, and provide an anticipated solution (Zio et al., 2008), (Pecht, 2010), (Dou et al., 2012).

In the framework of the Power-OM project (<http://power-om.eu/>), the research is focussed on machine tool performances. Within this project in order to improve the competitiveness of machine tools users and manufacturers, the aim is to use energy consumption monitoring and profiling, in an easy to implement Condition based Maintenance (CbM) technique, and manage it as a mechanism to improve the overall business effectiveness, under a triple perspective:

1. Optimizing maintenance strategies based on the prediction of potential failures and guiding the planning of maintenance operations: to schedule maintenance operations in convenient periods and avoid unexpected equipment failures.
2. Operation: Managing the energy as a production resource and reduce its consumption
3. Product reliability: Providing the machine tool builder with real data about the behaviour of the product and their critical components

In that sense, this paper presents an architecture where from CbM machine tool data, a set of added-value services are defined in order to track machine tool performances in terms of unit health, maintenance and energy consumption. This architecture allows failure anticipation through the local tracking of performances with real-time current analysis. Moreover a deeper analysis along the time is possible with the remote level allowing to provide a fleet-wide dimension.

This paper is composed of 4 sections, the first one which presents the introduction and context of this research. The second one presents the need of implementing CbM strategies and remote services to achieve the mentioned objectives. The third chapter presents an overview of the technical solution architecture for supporting the continuous monitoring, assessing machine health and for providing fleet-wide services. Finally the fourth chapter presents an overview of the first results obtained at the remote level.

2. PROBLEM STATEMENT

Condition based Maintenance (CbM) activities include obtaining data from sensors coming from the machine for its analysis which helps to measure, track and to understand the machine tool performances. This approach allows to compute some indicators at the local level to monitor the local health of a machine.

Classical machine tool monitoring techniques are related with acoustic and vibrations techniques (Abele et al., 2010). However, the different condition monitoring methods like vibration or acoustic monitoring usually require expensive sensors (Alzaga et al., 2014). A first originality to achieve a proactive condition monitoring approach, with non-intrusive monitoring techniques, affordable in terms of cost and effective, is to use the current analysis for assessing the health status of the machine. A special focus is paid to the critical components: the spindle and the linear axis. In that sense, recent research has been directed toward electrical monitoring of the motor with emphasis on inspecting the stator current of the motor (Kliman and Stein, 1992), (Benbouzid, 2000), (Alzaga et al., 2014). During the project it is expected to learn the relationship between electrical signal and the wear of the spindle and the linear axis. Based on the signature analysis results, the health index of these components could be computed and associated with different degradation modes of the components (e.g. gears missing tooth, etc.).

Moreover, to provide added-value information another originality of the approach is the use of a remote level. In that sense offline data from each machine is sent to an eMaintenance platform, which allows storing data from different machines. This level enables fleet performances management and monitoring. This dimension allows to benefit of more data and information allowing to make performances comparison across the fleet and along the time.

To achieve these objectives, in the following section an overview of the expected architecture with a local and a remote module is given.

3. TECHNICAL SOLUTION ARCHITECTURE

Within the research work, a technical architecture is needed for:

- Monitoring the health conditions of the machines and its main components (spindle and linear axis).
- Assessing operational conditions evolution in order to understand the usage of the machine and the environment. The characterization of operational conditions helps to understand the relationship between the health status and the usage of the machine (machine working for a long period of time with high loads, or high speed, etc). This way it is possible to have a better

understanding of the impact of the operational conditions on the evolution of the degradation modes (Medina-Oliva et al., 2013).

- Providing a fleet dimension in order to collect, gather, merge, summarize, share, compare and capitalize the data and knowledge coming from a set of units.

In order satisfy to these requirements, a technical architecture based on two levels is addressed. First, the local one, at the machine level, where proper CbM technology should be implemented. Second, at the remote side where relevant information coming from the local level is received, and, where having the fleet-wide information, the decisions and support can greatly improve operation and maintenance strategies.

Even if both levels aim at providing indicators about the health of a unit based on the monitored variables, a distinction between the roles of each level should be done. In that sense, the definition of local and remote services is provided as follow:

- Local level: this level is focused on real-time data processing. It is thus possible to make calculations and indicators directly from the machine for providing the current state of the unit as well as faults detection and prediction.
- Remote level: means to treat historical, offline data. And even if it's used for the machine/local level, the computation of the indicator or service is performed remotely. This level provides the indicators on the state of a machine or of a fleet (of machines or components). Moreover, data can be stored and comparisons about the evolution of indicators along the time could be also performed.

Following these requirements, the two levels are presented as follow.

3.1 CbM local module

The CbM module will collect two types of data:

- the fingerprint data use for motor current signature analysis. The fingerprint is the electrical signature of the machine in a specific time domain. This data is collected while the machine is running under a test procedure. Based on the electrical signature analysis, the relationship with failure modes are established and the health status/condition of the machine is assessed.
- operational data for inferring the use of the machine and for contextualization the component/machine performances.

To obtain this data, the CbM module should allow to monitor:

- CNC internal real time signals (i.e. jerk),
- External real time data (from external sensors of the machine)
- Other CNC data (for operational conditions)

In that sense, the machine should be equipped with a monitoring system. In the case of the Power-OM project two local modules could be used: the Genior Modular OA, Open Architecture, hardware from Artis, or modern CNC systems, like the CNC8070 from Fagor Automation where sensors can be already connected to the CNC or digital drive system. The CbM local module should

synchronize both flows of real-time data and have the possibility to combine different prediction algorithms to implement the multi-sensor fusion.

This module will allow the access to the machine internal data and the data from external sensors and then upload these data to the remote server.

3.2 Remote module: e-Maintenance platform

A remote server will enable to receive data coming from different machines in different sites (the fleet), aggregate and make them semantically comparable, considering the different context they have: Technical differences (the machines are not exactly the same), operational conditions, historical failures, etc.

The e-Maintenance solutions provide mechanism that supports organisations to transfer data to decision from a system perspective. Moreover decisions are based on the understanding of data relationships and patterns. Materialized as a set of interoperable, independent and loosely coupled information services a framework with its inherent infra-structure (i.e. e-Maintenance Cloud –eMC-) it is possible to provide fleet wide, continuous, coordinated service support and service delivery functions for operation and maintenance.

The e-Maintenance platform proposed for the Power-OM project is based on KASEM® (Knowledge and Advanced Service for E-Monitoring) (Léger, 2004). It is a collaborative e-Maintenance platform, integrating engineering, proactive maintenance, decision making and expertise tools. The foundation of KASEM is SOA, Services Oriented Architecture and Enterprise Service Bus: a software and systems architectural principles, based on Web services, to bring together a set of enterprise applications through XML-based engine.

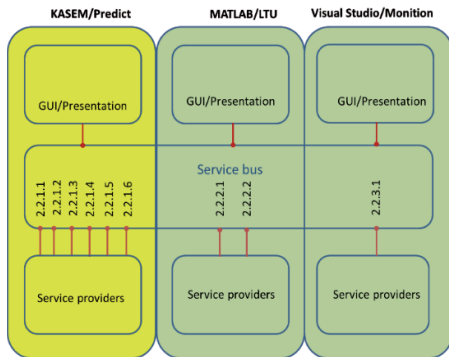


Figure 1. Infrastructure among user applications and service providers

The e-Maintenance platform support Power-OM infrastructure, in a true SOA way, through services and data exchanges via web services between the different modules in the Power-OM platform (Figure 1). This flexible infrastructure makes it possible for different partners to develop the different modules in different development environments.

This way, by means of the Power-OM platform, it should enable to:

- a. To collect and store data from the local level from the machines composing the fleet,
- b. To manage and store knowledge by means of a knowledge-based system,
- c. To share data from the units composing the fleet,
- d. To achieve large data analysis from stored, offline data from the fleet,
- e. To facilitate access, communication, user-friendly interpretation, and decision through presentation of remote services by means of web browser visualization.

in order to support the fleet-wide services by providing the corresponding information. This way, information from different machines could be uploaded and analyse in the Power-OM platform. An example of the typical fleet-wide architecture is provided in Figure 2.

3.3 Services for Machine Tool Performance Monitoring

The architecture presented in the last section set the basis to provide services at the local and at the remote level as explained as follow.

3.3.1 Local health monitoring through Current Signature Analysis (Fingerprints)

In Power-OM project the concept of machine fingerprint is used. It is the electrical signature of the machine in a specific time domain without operation influences.

Fingerprint raw data are processed and the relevant signal features are extracted. These relevant features are used to compare the fingerprints through its values on time and that will give the health assessment of the machine. To compute the health index of the machine it is necessary to analyse the relationship between the current and other sensors with the health status. The idea consist on the correlating any load and speed variation with current and voltage variations to reflect loads, stresses and wear of components. But to do it is necessary to identify the healthy electrical signature of the machine, which is called the reference fingerprint, and the degraded ones associated to the different failure modes.

However, to do this in an industrial environment, first it is necessary to learn from experiments on test benches the potential correlation between electrical signals and wear, stresses and load on the machine. In the Power-OM project different test benches are used: Gearbox test bench, spindle test bench and linear guides test bench, as explained in (Alzaga et al., 2014).

Once fingerprint experimental phase is undergone in test benches, the main signals, features and the relationship with failure modes are established, the implementation will continue with an operational milling machine from Goratu using the Genior Modular OA, Open Architecture, hardware from Artis.

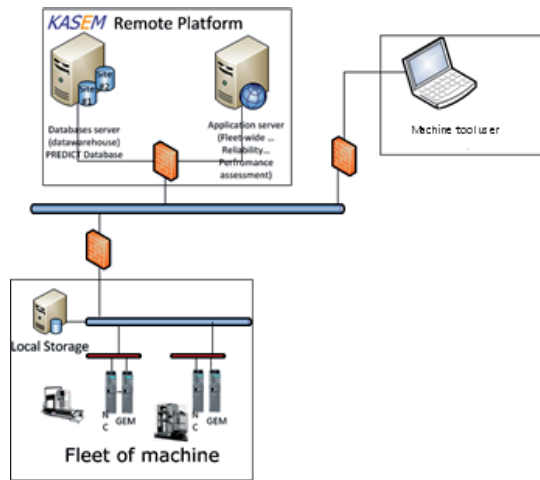


Figure 2. Example of typical architecture including a fleet of machine and machine tool end-user

With the data obtained from the fingerprint test, different features are computed such as the mean, median, variance, etc. on time domain and frequency analysis, as well as the health index. These results will be then sent to the remote level. This way, different types of degradations could be early detected such as gears degradations (e.g. missing tooth, chipped tooth, etc.).

3.3.2 Remote fleet-wide services

The remote level provides a set of possibilities and added-value services as described as follow.

3.3.2.1 Remote fleet-wide health monitoring

The results of the machine health assessed locally, are then sent to the remote level. This way added-value services could be performed from a fleet perspective.

The health assessment provides a global overview of the health state at different abstraction levels (Medina-Oliva et al., 2013). To do it, it is necessary to consolidate and aggregate data coming from machines in order to assess indices reflecting the operation condition of a set of machines. Based on the aggregated and synthesized health index, this service will allow to point out the degradation of component within the fleet.

The health assessment with a fleet perspective could allow to:

- To provide synthesized indicators representatives of the health of the fleet in order to have a global view of its state to achieve the expected missions.
- Based on the state of the fleet/plant/machine, to see if it's necessary to perform some "improvement actions" (e.g. maintenance actions) at a lower level (e.g. component).
- To follow-up the evolution of the health index along the time

- To compare the health assessment between the different sub-fleets to identify priority actions based on the critical health assessment.
- To compare performances of parts coming from different providers
- To contextualize the health indicators according to the operational conditions of the machine, allowing a better understanding of the use of the machine.

For providing the necessary synthetic view of the health performances at different abstraction levels data, health index data from the components need to be consolidated and aggregated (Figure 3). This way, "individual" health indices (e.g. spindle health indices) will be merged and aggregated within the fleet dimension allowing to provide a global vision of the health state of a fleet along the time. The approach is summarized in Figure 3.

3.3.2.2 Remote fleet-wide power assessment and monitoring

Based on the data gathered as operational data, it is possible to provide fleet-wide services to compute energy efficiency indices at different abstraction levels. The energy usage and efficiency assessment will thus be made on relevant time basis including operation condition of the machine for assessing efficiency.

This way, the basis for power optimization of machine tools will be provided. In that sense, it will possible to minimize for example the power consumption of the machine while minimizing the use of energy by putting subsystems like cooling fluid, hydraulic pumps, cooling fans etc. in idle/sleep mode (Anderberg et al., 2012).

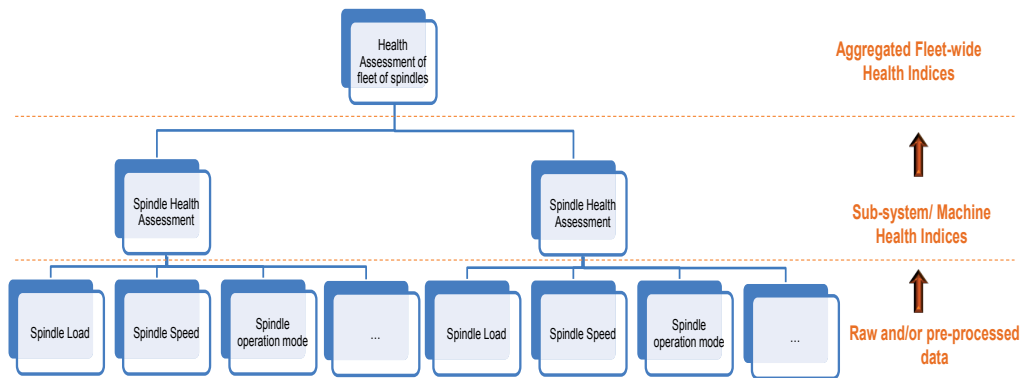


Figure 3. Example of health assessment indices at different abstraction levels

3.3.2.3 Remote fleet-wide prognostic assessment

The estimation of the Remaining Useful Life (RUL) allows to warn the time to wait to reach the minimum performance threshold. In that sense, RUL can integrate information about the power consumption as well as operational data in order to estimate the appropriate replacement times.

Fingerprints and collected operational data provide the contextual information needed to understand and predict the different machines and components condition and RUL.

From the fingerprint (current analysis), it is possible to track the variations in the features. Moreover, with data mining techniques it is possible to discover patterns and interesting relations between variables in the operational context data. These patterns and relations is the 'Knowledge' of how the use of the machine will affect the fingerprint and thus the state of the machine/component.

Moreover, by use of advanced algorithms it is possible to track the variations in fingerprint features in a multidimensional space and estimate the time to reach unaccepted areas (fault). This estimated time will be the RUL for the machine/component. The borders to those unaccepted areas will continuously be adjusted through the knowledge of the history of the machine/component.

4. FIRST RESULTS AT THE REMOTE LEVEL

A start of the first remote level services have been computed with the operational data that have been collected locally and sent to the remote server for a further fleet health assessment analysis.

Regarding the connectivity issues two scenarios have been taken into account:

- Machine directly connected to the internet, using Ethernet or GSM, and where the files can be uploaded under VPN, HTTP or FTP.
- Machine not directly connected and where files can be manually uploaded or sent by e-mail.

In the remote level, offline data was stored in KASEM®. This level allows the follow-up short and mid-term performances for deeper analysis of the fleet performances in order to bring relevant information for decision makers.

Some examples of analysis of operational data are provided in Figure 4 and Figure 5. These figures shows the characterisation of the spindle speed. Figure 4, shows a pie chart illustrating the working time of the machine based on different categories of spindle rotation speed (working speed).

Figure 5 allows to complete the analysis with a box plot graph. This kind of representation allows to show different statistical information concerning the variability of data. The bottom and top of the box are always the first and third quartiles. The dashed line inside the box is the mean and the full line is the median. Moreover, the ends of the whiskers represent the percentiles 5 and 95. This information allows one to see data variability for each working speed.

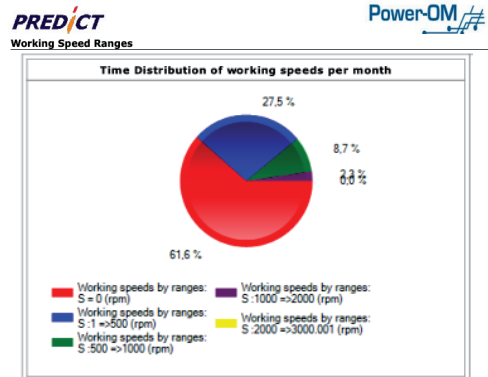


Figure 4. Pie chart showing the time distribution of the working speed ranges per month

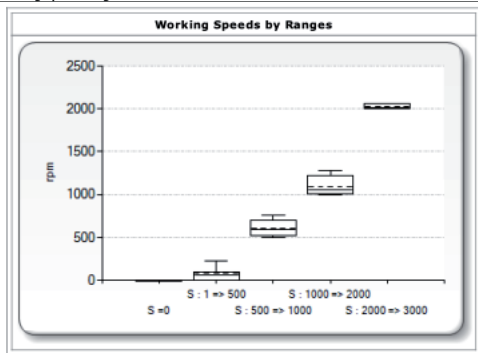


Figure 5. Box plot showing the working speed ranges variability

5. CONCLUSIONS AND PERSPECTIVES

This paper presents the results and the possibilities of an architecture allowing to monitor machine tool performances in order to support a proactive degradation detection through the analysis of the current signal. To do it, two levels of services are provided:

- Local level: at the machine level, where CbM strategy should be implemented. This level is focused on reliable real-time data processing. It is thus possible to make calculations and indicators directly from the machine for providing the current state of the unit as well as faults detection and prediction.
- Remote level: where historical, offline data coming from the local level is received, processed and managed with a fleet-wide approach, allowing to support and improve the decision making process. This level provides the indicators on the state of a machine or of a fleet (of machines or components). Moreover, data can be stored and comparisons about the evolution of indicators along the time could be also performed.

Moreover, the remote level provides a number of possibilities to support a proactive maintenance approach, such as the health assessment at the fleet level, the assessment and the optimization of power consumption and the assessment of the RUL. This remote dimension brings added value information thanks to the ability of the e-Maintenance platform, to manage, process and present the data coming from each unit within a fleet dimension.

This way it will be possible to implement and to show the added-value of the Power-OM solution for providing new services locally and remotely. Those generic added-value services could be implemented in different kind of machines according to the availability of data.

To show the feasibility of the proposed approach, the first results obtained at the remote level were shown in section 4. These results allow to illustrate the first steps towards the proposed services. Furthermore it was possible to validate the first steps towards the Power-OM architecture allowing to:

- To obtain data from the local CbM of a fleet of machines through communication mechanisms between machine and the remote server
- To interoperate and exchange data between the local and remote level
- To manipulate data from a fleet of units to provide knowledge sharing and capitalization.

Further service developments are required in order to implement the described services at the local and remote level. Moreover, a validation phase should be considered in order to verify that the services address the expectations of the machine tool end-users.

6. ACKNOWLEDGEMENTS

The work presented in this paper is part of the research work still in progress in the project Power-OM, funded under FP7 under grant agreement no. 314548, as part of the Factories of the Future initiative.

7. REFERENCES

- [1] Abele E., Altintas Y., Brecher C. (2010). Machine tool spindle units, CIRP Annals - Manufacturing Technology, 59/2, 781-802.
- [2] Alzaga A., Konde E., Bravo I., Arana R., Prado A., Yurre C., Monnin M., Medina-Oliva G. (2014). New technologies to optimize reliability, operation and maintenance in the use of Machine-Tools. Euro-Maintenance conference, 5-8 May 2014, Helsinki, Finland.
- [3] Anderberg, S., Beno, T., Pejryd, L. (2012) Energy and Cost Efficiency in CNC Machining from a Process Planning Perspective, Sustainable Manufacturing, pp 393-398.
- [4] Benbouzid M. E. H. (2000). A review of intrusion motors signature analysis as a medium for faults detection. IEEE Transactions on Industrial Electronics, 47(5).
- [5] Dou D., Yang J., Liu J. and Zhao Y. (2012). A rule-based intelligent method for fault diagnosis of rotating machinery. Knowledge-Based Systems, vol 36, Pages 1-8.
- [6] Jung, B., Levrat, E., Crespo Marquez, A. and Erbe, H. (2009). Conceptual framework for e-Maintenance: illustration by e-Maintenance technologies and platforms, Annual Reviews in Control, PP. 220-229.
- [7] Kliman G. B., Stein J. (1992), "Methods of motor current signature analysis," Electric Machines and Power Systems. Vol. 20, no. 5, pp. 463-474.
- [8] Léger J-B. (2004). A case study of remote diagnosis and e-maintenance information system, Keynote speech of IMS'2004, International Conference on Intelligent Maintenance Systems, Arles, France.
- [9] Medina-Oliva G., Peysson F., Voisin A., Monnin M., Léger J-B. (2013). Ships and marine diesel engines fleet-wide predictive diagnostic based on ontology, improvement feedback loop and continuous analytics. Proceedings of 25th International Congress on Condition Monitoring and Diagnostics Engineering Management, 11-13 June, 2013, Helsinki, Finland.
- [10] Pecht, M. (2010). A prognostics and health management roadmap for information and electronics-rich systems, IEICE Fundamentals Review, vol. 3, no. 4, pp. 25 - 32.
- [11] Zio E., Baraldi P., Gola G. (2008). Feature-based classifier ensembles for diagnosing multiple faults in rotating machinery, Applied Soft Computing Journal 8 (4) 1365-1380.

PROCEDURE FOR RUL ESTIMATION IN INDUSTRIAL ASSETS

Angel Hernandez

Diego Galar

Numan Perales

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

+46 722 121 931

+46 725 390 810

+46 702 084 064

angel.hernandez@ltu.se

diego.galar@ltu.se

numan.perales@ltu.se

ABSTRACT

Today, prognosis is recognized as a key element of maintenance. However, the implementation of an efficient prognosis tool can be complicated in industrial and academic sectors, when we speak about academic sector, we refer to the research centers at universities who study the progress and new technologies related to prognosis, these centers are very important as they help the improve maintenance management in the industries. Since it is difficult to create effective models for different industrial assets. In this context, our general objective is to propose a procedure for implementing prognosis, from selecting the system or component to be analyzed to obtaining the estimation of the remaining useful life. We also explain different approaches to forecasting to estimate remaining useful life, the main objective of prognosis.

Keywords

Prognosis, maintenance, remaining useful life, system or component.

1. INTRODUCTION

The maintenance of all types of assets is increasingly important in the industrial and academic sectors. In recent years, industry has been concerned with improving maintenance methods, tools and techniques to achieve increased reliability, availability and security of equipment or assets, thereby reducing the high costs due to maintenance. But no matter how well these teams do, assets deteriorate over time and with stress or loads; as a result, maintenance activities are performed more frequently. Hence, industries and research centres have been interested in finding a way to predict when a failure can occur, given the current situation and operating profile. To achieve this, prognosis is performed; essentially, prognosis predicts the remaining useful life (RUL) of a system or asset before a failure occurs.

The prognosis is born from that in recent years the industry has been implementing maintenance corrective and preventive, where the result of applying this type of maintenance has not been very satisfactory due to the high cost that is generated, as to the fact corrective expect equipment failure can cause excessive cost because there is a risk that the component failure affecting other components, on the other hand, preventive maintenance generates high cost, as there are that do maintenance tasks with more frequency as stated earlier, it is for this reason that in the actuality is looking apply prognosis in order to improve both the management and reduce maintenance costs. Note that prognosis is a new technique that is having a very important role in the

maintenance world. Prognosis is based primarily on mathematical models using four approaches: model-based, data-based, experience-based and a hybrid model combining approaches and models [1]. The choice of technique depends on the available tools: the knowledge of data, dynamic and complex system, application requirements, available monitoring devices, amongst others [2].

The aim of this work is to develop a clear and precise process to make a prognosis following a series of steps, starting from the monitoring system selection to obtaining the estimate of the remaining useful life, so that maintainers can make decisions whether to perform an activity in the monitored system maintenance.

Many authors have written about prognosis, [Jardine, A], wrote about, review on machinery diagnostics and prognostics Implementing condition - based maintenance [Dragomir, O], wrote about, review of prognostic problem in condition - based maintenance. These are two examples of authors who have had interested in writing about prognosis, others be in the reference.

2. CONCEPT OF PROGNOSIS

Prognosis is a relatively new technique that has become an important maintenance system function, according to the International Organization for Standardization: "The forecast is estimating time to failure and the risk for one or more failure modes existing and future" (ISO, 133811 (2004)). Prognosis is also called the "prediction of the lifetime of a system," as it aims to predict the remaining useful life (RUL) before a failure, given the current state of the machine and operating profile [3]. This estimate is possible because machines usually go through a measurable signal of failure before failure occurs; this signal is used for the estimate. The observed condition can be determined by physical characteristics or process failures. For example, vibration analysis and oil analysis have been successfully used to monitor the presence of a computer failure.[4] Other parameters of alternative conditions that can be used in the prognosis data are noise, temperature, pressure, humidity, climate or environment data, among others.

Currently, the prognosis process takes numerous approaches ranging from failure rate models using historical data to physics-based models. Approaches are model-based, based on data and based on expertise [5]. Two salient features of prognosis are:

- Prognosis is mostly a process of prediction (future situation likely to occur).

- Prognosis is based on the notion of failure, which means that it is associated with a degree of acceptability (the predicted position with respect to a reference to be assessed).

Both features are distinguished by the clarity of presentation, yet are bound together in reality [2]. Figure 1 shows prognosis in maintenance

3. APPROACHES TO PROGNOSIS

As discussed above, there are three main types of approaches.

3.1 Model-based approach

A model-based approach is applicable in situations where an accurate mathematical model can be constructed from the first principle of the failure modes of a system. Therefore, it requires specific knowledge of the mechanical theory and the monitored equipment [6]. An example of this approach is based on fatigue, where fatigue represents the initiation and propagation model structural abnormalities. Moreover, these methods typically use waste as characteristics and statistical techniques to define the thresholds that allow the detection of faults; these residues are the results of consistency checks between measurements obtained from a real system and the results of a mathematical model. The premise is that the residuals are larger in the presence of a malfunction, and smaller in the presence of normal disturbances [1].

The main advantage of this approach is the ability to incorporate physical understanding of the monitored system. In addition, if the understanding of the degradation improves, the model can be adapted to improve accuracy and to address problems and subtle performance issues. The main disadvantage of this approach is that it may be difficult or even impossible to achieve system performance; thus, this type of approach cannot be applied to very complex systems [7]. Techniques used in this type of approach include: the particulate filter model based on physical variables (fatigue) and the autoregressive moving average model (ARMA).

3.2 Approach data

The data-driven approach uses real data and on-line data collected by sensors or by operators. Such data can be approximated; therefore, we can keep track of the features that reveal the degradation of the components, allowing us to forecast the overall system behaviour. In fact, in many cases, the data (both input and output) are the main source of a deeper understanding of system degradation. Forecasting is performed using indicators of degradation and statistical techniques or artificial intelligence methods. Statistical methods include: multivariate linear and quadratic discriminators, partial least squares, among others. Techniques of artificial intelligence include: neural networks, fuzzy systems, decision trees, among others [8].

The main advantage of this approach is its ability to transform the noisy high-dimensional data into lower dimensional information for prognosis and decision-making. Artificial intelligence techniques are increasingly applied to calculate the estimated useful life remaining and show better performance than conventional approaches. In practice, however, it is not easy to apply artificial intelligence techniques; its main disadvantage is

the lack of effective procedures to obtain training data and expertise. To this point, most applications have used experimental data for model training. Briefly stated, data based approaches are highly dependent upon the quantity and quality of the system's operating data [1].

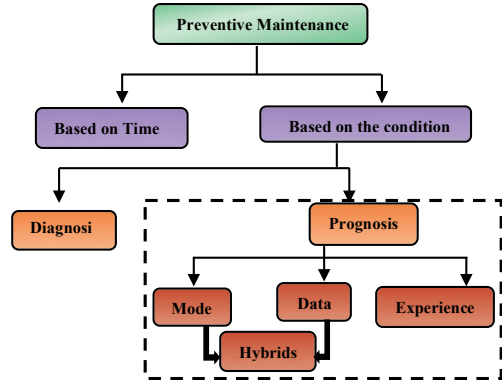


Figure 1. Prognosis in maintenance

3.3 Based on experience

Experience-based prognosis is less complex and requires fault history data of equipment or component design recommendations under similar operation. Therefore, experience-based prediction is applicable to mass production equipment. This prognosis is based on probabilistic models of degradation, where the failure and / or inspection of the data are compiled from legacy systems and a Weibull distribution or another statistical distribution that fits the data. An example is scheduling the maintenance of a component of low criticality, with little or no associated parameters. In this case, the prognosis for a component to fail or degrade to an unacceptable condition should be based solely on the analysis of past experience. Depending on the complexity of maintenance and criticality associated with the component, the prognosis system can be configured for a range of maintenance then updated as more data become available. Having an automatic database maintenance is important for the implementation of prognosis based on experience [9].

The main advantage of this type of approach is that it is easy to implement, as long as sufficient condition and significant data are available. Its main disadvantage is that, the derived reliability functions are specific to each set of assets and not for one specific, this means that the RUL estimated is for the whole set of assets, thus generating weak results in the precision estimation of RUL.

4. STEPS TO PERFORM PROGNOSIS

This section explains a step by step process to make a prognosis, from the selection of equipment or system to monitor to obtaining the remaining useful life. Figure 2 provides an outline of the process.

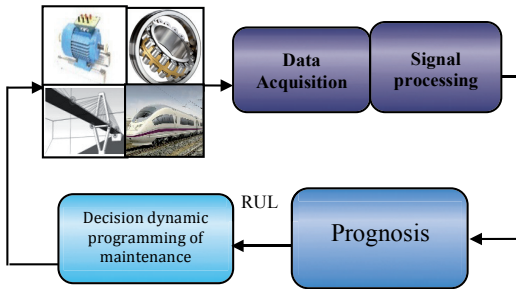


Figure 2. Steps to achieve prognosis

4.1 Data acquisition and signal processing

When we talk about industrial assets, we are referring to the components, machines and systems that can be monitored. Components include bearings, gears, and others. Machines include rotating machines (pumps, turbines, compressors), electrical machinery (motors, transformers) among others. Finally, a system consists of a series of components or machines, for example: a plane, a nuclear plant, an automotive structures (bridges, buildings, roads) among others.

Once the asset to monitor is determined, we proceed to data acquisition. To do this, we must select the most appropriate monitoring technique depending on the equipment or system to be monitored. Monitoring techniques include vibration analysis, oil analysis, thermography, noise analysis, among others. Using these techniques, we can collect data from physical variables (temperature, vibration, viscosity, voltage, sound waves, among others). We obtain the dataset through multiple sensors placed in specific locations on our assets. These sensors collect information, which is then processed and stored using specialized software.

Some sensors commonly used for these purposes are the following:

1. Vibration sensors: Transducers or contact speed sensors, laser vibrometers, accelerometers, force transducer.
2. Temperature sensors: liquid bulb thermometer, thermocouples, thermistors, semiconductors.
3. Sensors sound: microphones.

Once we have the data, we must perform signal processing to eliminate any contamination from signals. In most cases, data are contaminated by signals from the environment where the computer is located, causing the data to be inconsistent; an example of such a signal is noise pollution. To achieve this goal, we can use statistical signal processing algorithms.

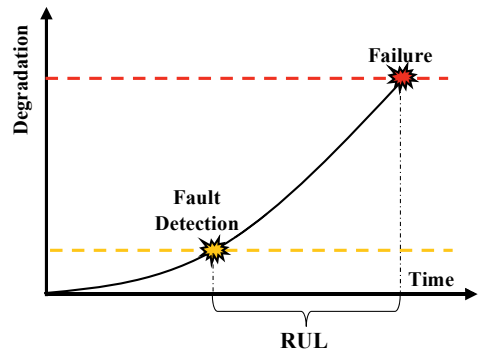
In this framework, all the historical data sets in the computer software will be extracted to new database degradation. This database degradation can be used to build new knowledge to generate the degradation threshold. The threshold can be used as a model of behaviour to identify the presence of failure in a new

data set of equipment operation. The thresholds can be formulated by implementing statistical techniques or artificial learning techniques in the existing historical data sets. Usually processing a low level signal is good enough to obtain a satisfactory threshold [10]. The most important part is that the threshold must consistently decide the level, i.e. whether it is acceptable or unacceptable in any pattern observed in the operating parameters. The new set of observed data can be evaluated with the behavioural threshold for the presence of failure. If the level is within the control limit threshold, the equipment indicates normal behaviour. Therefore, the equipment may be available for work without any maintenance action being required. If, on the contrary, the level exceeds the control limit threshold, this means a fault has been observed in the data, and we must estimate the remaining life of the equipment.

4.2 4.2 Prognosis

The data acquired for the equipment or system being monitored will show any degradation of the system, and if this is the case, the next step is prognosis. As stated earlier, prognosis is based on estimating the remaining useful life, or the time from when the system begins to degrade until it fails completely; see Figure 3

The remaining useful life can be estimated using the techniques discussed above, depending on the type of asset being monitored. In the case of components, the approach is based on models, as both the theory and the design are known. When data are collected by sensors, the data-based approach should be used. Finally, when



a team cannot collect enough data and have only a fault history, the experience-based approach is applicable [11].

Figure 3. Remaining useful life.

4.3 Decision making (maintenance schedule).

After obtaining the estimate of the remaining useful life of the asset or system, maintenance engineers must make a series of decisions and plan maintenance activities using the maintenance team, as well as maintenance resources, such as spare parts. These decisions centre on whether to perform a maintenance activity or not; it is important to make a good estimate to minimize maintenance costs. At times, this may be appropriate when the asset is still functioning properly.

5. EXAMPLE OF A PROGNOSIS

As an example, we discuss the propagation of cracks in rotating machinery using a model-based technique and the particle filter approach. Figure 4 displays a rotary machine and crack propagation.

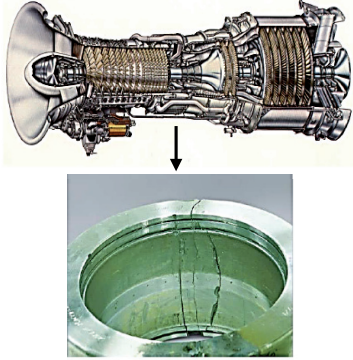


Figure 4. Crack Propagation in Rotating Machinery.[7]

Using a numerical approach, the posterior probability density function (PDF) is given as:

$$p(X_k | Z_{0:k}) \quad (1)$$

From the data & information, we have:

- Historical trajectory of degradation. $\{Z_1, Z_2, \dots, Z_N\}$ (2)

- Degradation model.
 - Hidden Markov Process. $X^i_k = f_k(X_{k-1}, \omega_{k-1})$ (3)

- Measurement equation. $Z_k = h_k(X_k, \nu_k)$ (4)

Hypothesis:

- System Model.
 - X = hidden vector state of degradation.
 - ω = random process noise vector.
 - f = vector of nonlinear dynamic function.
 - k = time step index.
- Measurement equation.
 - ν = noise vector.
 - h = as nonlinear, vector function.

Once the hypothesis is determined, we can derive the equations as follows:

- Prediction of Monte Carlo, state trajectories N (= particles).

$$X^i_k \quad i = 1, \dots, N \quad (5)$$

- Chance of Observation (weight of particles).

$$p(Z_k | X^i_k) \quad (6)$$

The distribution of the system's state is given by the estimation of the state and the distribution of time to failure.

- State estimation is given as

$$p(X_k | Z_{1:k}) = \sum_{i=1}^N w_k^i \delta(X_k - X^i_k) \quad (7)$$

- Failure time distribution is given as

$$p(t_f | z_{1:k}) = \sum_{i=1}^N w_k^i \delta(t_f - t^i) \quad (8)$$

Once the equations are derived, the data are entered; in this case, we use sample data as follows:

- Number of particles: 5000
- Five measurements at time intervals: $k_1 = 100, k_2 = 200, k_3 = 300, k_4 = 400, k_5 = 400$.
- ω k, k : Gaussian noise.
- d^* : 80

After the data are entered into the equations, the equations are used to generate a graph showing remaining useful life of the part being analyzed. At this point, the maintenance engineer should compare the result with the remaining life which has been calculated in the past and make a decision whether or not to replace the part; as mentioned above, over time the system will degrade; in this case the crack will get bigger. Figure 5 shows the graph of the RUL obtained for this example.

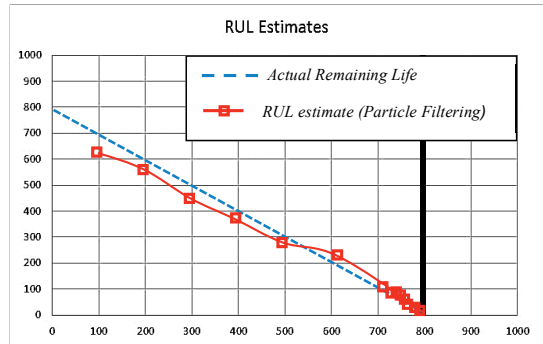


Figure 5. RUL estimate.

6. CONCLUSION

Sustainable development is the integration of economic strategies with social and environmental concerns to optimize an industrial process. Therefore, maintenance is becoming increasingly important in the planning and development of industries once. Concepts such as corrective and preventive maintenance are currently being replaced by predictive maintenance using a new technique called prediction. Today the forecast is recognized as a key element in strategies for maintenance, but due to lack of knowledge about how it works, is not yet fully established. This

paper offers four main steps to apply for a prognosis (data acquisition, filtering signals, prognosis and decision making). Note that the forecast is used to help predict failure so the ultimate responsibility for deciding when to perform a maintenance activity remains in the hands of maintenance engineers.

7. REFERENCES

- [1] Dragomir, O. Gouriveau, R. Dragomir F. Minca, E. Zerhouni, N. "review of prognostic problem in condition – based maintenance", version 1-21, sep 2009
- [2] Medjaher, K. Gouriveau, R. Zerhouni N. "A Procedure for Failure Prognostic in Dynamic Systems" version 1 - 10 Jun 2009,
- [3] Jardine, A. Lin, D. Banjevic, D " A review on machinery diagnostics and prognostics implementing condition – based maintenance", Vol. 20, pp. 1483- 1510, 2006,
- [4] Lee, J. Djurdjanovic, D. Qiu, H. Liao, H. "Intelligent prognostics tools and e-maintenance", 57(6): p. 476-489. 2006.
- [5] Tran, V. Yang, B. Oh, M. Tan, A. "Machine condition prognosis based on regression trees and one-step-ahead prediction, 2006.
- [6] Muller, A. Suhner, M. lung, B "Formalisation of a new prognosis model for supporting proactive maintenance implementation on industrial system".
- [7] Chelidze, D "Multimode damage tracking and failure prognosis in electromechanical system", Vol. 4733, pp. 1-12, 2002.
- [8] Wang, W. Golnaraghi, M. Ismail, F. "Prognosis of machine health condition using neuro-fuzzy systems", Vol. 18, pp. 813-831, 2004.
- [9] Byington, S. Roemer, M. Galie, T. "Prognosis Enhancements to Diagnostic System for Improved Condition Based Maintenance". 2002
- [10] Wang, W "A two-stage prognosis model in condition based maintenance". 182(3): p. 1177-1187, 2007.
- [11] Butler, S. "Prognostic Algorithms for Condition Monitoring and Remaining Useful Life Estimation" A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy, September 2012, 243 pp.

eMaintenance 4B

Multi-Criteria Data Quality Assessment

Maintenance perspective

Mustafa Aljumaili

Division of Operation, Maintenance
and Acoustics
Luleå University of Technology
Luleå, SE-971 87, Sweden
+46 920 49 15 53

mustafa.aljumaili@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

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

ABSTRACT

Data quality (DQ) in maintenance has become an increasingly important aspect to many firms as most of the maintenance planning and implementations are based on data analysis. Poor DQ has adverse effects at the operational, tactical, and strategic levels of any organization. Respectively, poor DQ reduces customer satisfaction, leading to poor decision making, and has negative impacts on strategy execution. To improve DQ as well as to evaluate the current status, DQ need to be measured following the fact that only what can be measured can be improved. A measure for DQ could be an important support for decision makers. In order to assess DQ, related attributes should be defined. These attributes could be related to the data itself, to the metadata, or to the data representation schemes. After defining these attributes, an assessment model should be used to evaluate these attributes. The purpose of this paper is to propose a model for DQ assessment. Therefore, a study of DQ attributes and the possible metrics that could be used to measure these attributes was undertaken. The proposed model will be applied on dataset provided by the Swedish Transport Administration (Trafikverket) for validation and to find an estimation measure of the DQ.

Keywords

Data Quality, Information Quality, Maintenance, eMaintenance, Attributes, Metrics, Assessment.

1. INTRODUCTION

The subject of information and DQ has been receiving significant attention in the recent years in many areas, including; communication, business processes, personal computing, health care, and databases. As a result of today's world of massive electronic data sets and difficult policy decisions, DQ problems can create significant economic and political inefficiencies [1]. Therefore, information is critical for every aspect of modern life and its quality largely determines the quality of decisions made, ultimately it affects the quality of activity and action outcomes in organizations and in the society in general. Data is the foundation for information and it is one of the most important criteria for making strategic business decisions within organizations.

While there are no universally agreed definitions of DQ, there is no dispute about the importance of it and the consequences. Inadequate DQ has major financial consequences and led to incalculable losses [2]. It has been noticed that, U.S. businesses pay \$600 billion a year due to a lack of DQ [3]. In addition, about 75 per cent of organizations have identified costs originating from dirty data, i.e. poor quality [4].

In particular, “good” or “improved” DQ is often not an end in itself, because it is the quality of the decisions, not the data that ultimately matters [1]. In general, quality of data is influenced by three main factors: the perception of the user, the data itself, and the process of accessing the data. The three factors can be seen as the subject, object, and predicate of a query. Each factor can be a source for Information Quality (IQ) metadata [5]. A means to assess IQ for decision-making is vital. Without clearly defined attributes and their relationships, we are unable to assess IQ and may be unaware of the problem [6].

In order to improve DQ as well as to evaluate the current status, DQ need to be measured. In literature, several authors pointed out that: “Only what can be measured can be improved” [7] [8] [9]. Therefore, the need for DQ measurement approach is important to determine the level of DQ over time. In fact, many companies are running DQ initiatives in an attempt to improve DQ. But how many of them really measure the quality of their data? What is the suitable assessment method and how many of them really know how to implement a measurement system?

In this study, an attempt to define a DQ assessment method was undertaken. Therefore, a study of the attributes and metrics of data is presented. Ranking these attributes and finding the evaluation for each is discussed. The rest of the paper will be organized as the following: the next section will discuss related work. Section three will describe the DQ assessment process. Section 4 and 5 discuss the attributes and metrics while 7, 8 and 9 will present the proposed model, validation of model, results and conclusions.

2. RELATED WORK

The subject of DQ attributes has been studied extensively in literature. However, the assessment of DQ needs more emphasis. Despite a decade of research and practice, only ad hoc techniques

are available for measuring, analyzing, and improving IQ in organizations [10].

In 1998, Richard Y. Wang has developed a TDQM (Total Data Quality Management) methodology. The idea of TDQM is to deliver high quality information products (IP) to information consumers. It aims to facilitate the implementation of an organization's overall DQ policy formally expressed by top management. This method was based on the field of product manufacturing which has an extensive body of literature on Total Quality Management (TQM) that deals with information as a product [11].

In 2001, Yang W. Lee et al have developed a methodology, called AIMQ to form a basis for IQ assessment and benchmarking. The methodology was illustrated through its application to five major organizations. Their methodology encompasses a model of IQ, a questionnaire to measure IQ, and analysis techniques for interpreting the IQ measures [10].

In 2004, Enrique Herrera-Viedma et al suggested an evaluation methodology based on fuzzy computing with words aimed to measure the IQ of Web sites. This methodology included both qualitative and user oriented. This study mainly composed of two main components, an evaluation scheme to analyze the IQ of Web sites and a measurement method to generate the linguistic recommendations. The evaluation scheme was based on both technical criteria related to the Web site structure and criteria related to the content of information on the Web sites.

Besiki Stvilia et al in 2007 have proposed a general IQ assessment framework. Unlike the context-specific IQ assessment models, which usually focus on a few variables determined by local needs, the suggested framework consists of comprehensive typologies of IQ problems, related activities, and taxonomy of IQ dimensions organized in a systematic way based on sound theories and practices.

Another study has been conducted by Mona Alkhattabi et al, 2011. The focus of this study was on assessing IQ in e-learning systems. The proposed framework consists of 14 quality dimensions grouped in three quality factors: intrinsic, contextual representation and accessibility. They have used the relative importance as a parameter in a linear equation for the measurement scheme. The data collection and evaluation processes were automated using a web data extraction technique and results on a case study were discussed. [12].

Finally, Hongwei Zhu and Harris Wu in (2011) have developed a set of metrics and a framework for assessing data standard quality. The developed metrics included completeness, relevancy, and a combined measure. They tried to evaluate the framework on a data standard for financial reporting in United States, the Generally Accepted Accounting Principles (GAAP) Taxonomy encoded in eXtensible Business Reporting Language (XBRL), and the financial statements created using the standard by public companies. Their results have shown that the data standard quality framework is useful and effective [13].

3. DQ ASSESSMENT PROCESS

The aspect of product manufacturing has an extensive part of literature on Total Quality Management (TQM) with principles, guidelines, and techniques for product quality. Based on TQM, knowledge has been created for DQ and IQ management.

Therefore, an organization would follow certain guidelines to build DQ project, identify critical issues, and develop procedures and metrics for continuous analysis and improvement. But we should mention here that these approaches have limitations [14].

These limitations arise from the nature of data and information production process and the use of this information. Data can be used by different consumers in different contexts.

Defining, measuring, analyzing, and improving IQ continuously is essential to ensure high quality data. In general, the DQ assessment process consists of several processes that should be applied by an organization, the users and the developer as well. These steps can be described in the following (see Figure 1):

- 1) The definition step identifies important DQ dimensions and the corresponding requirements of DQ.
- 2) The measurement step defines and produces the DQ metrics and measures necessary to evaluate DQ.
- 3) Analysis step identifies root causes for DQ problems and calculates the impacts of poor quality information.
- 4) Improvement process provides suitable techniques for improving DQ.

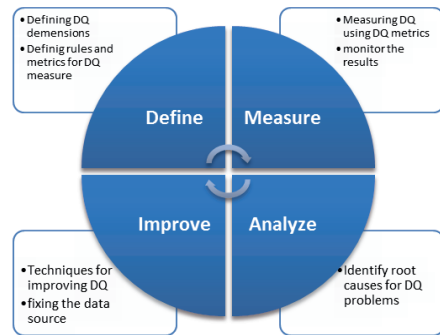


Figure 1. Data Quality Assessment Process, based on [1]

All of these steps should be applied along DQ dimensions according to requirements specified by the consumer. Therefore, the context plays important role in each step of this process.

4. DIMENSIONS AND ATTRIBUTES OF DATA QUALITY

Study of DQ dimensions is necessary for any organization in any business for many reasons. First, it helps facilitate the assessment and measurement of the quality of the data. Second, it provides a framework for guideline and improvement plans of DQ. When developing these measures, the company must determine what is to be measured, what set of DQ dimensions that are important to its mission and operations. Many of the dimensions are multivariate in nature. Therefore, the attributes which are important to the firm must be clearly identified and defined [15].

Therefore, during the 1970s and 1980s, the Financial Accounting Standards Board (FASB) issued several concepts statements to guide the development of accounting and reporting principles for

use by U.S. companies. The key characteristics Statement No. 2 identifies and discusses are [16]:

1. Benefits of the information disclosure should exceed cost.
2. The information should be relevant.
3. The information should be reliable.
4. The information should be comparable.
5. The information should be material.

Taylor & Voigt (1986) has identified five kinds of values (i.e., dimensions) that IQ may possess: accuracy, comprehensiveness, currency, reliability, and validity [17]. Another intuitively derived classification was obtained through empirical studies engaged participants directly by selecting attributes that were important in their individual perceptions of IQ. Wang and Strong's (1996) study, for example, surveyed 137 users, yielding 179 different quality attributes that eventually reduced to 20 dimensions, and then further reduced to four primary IQ "categories [8]. Lee et al. (2002) have gathered IQ attributes from 15 studies, differentiating between those studies employing attributes from academic and practitioner points of view. They adapted the categories proposed by Wang and Strong (1996) and reduced IQ attributes to four main categories [10]. In a more recent review, Knight and Burn (2005) compared 12 earlier studies that used a variety of IQ attributes, reducing the number of attributes to 20 based on the frequency with which each attribute appeared across all of the studies examined [18].

As a result to all these studies in literature, DQ attributes can be summarized to the attributes in Table 1.

Table 1. DQ Dimensions [2]

Dimension	Definition
Accuracy	extent to which data are correct, reliable and certified free of error
Consistency	extent to which information is presented in the same format and compatible with previous data
Security	extent to which access to information is restricted appropriately to maintain its security
Timeliness	extent to which the information is sufficiently up-to-date for the task at hand
Completeness	information is not missing and is of sufficient breadth and depth for the task at hand
Concise	extent to which information is compactly represented
Reliability	extent to which information is correct and reliable
Accessibility	extent to which information is available, or easily and quickly retrievable
Availability	extent to which information is physically accessible
Objectivity	extent to which information is unbiased, unprejudiced and impartial
Relevancy	extent to which information is applicable and

	helpful for the task at hand
Usability	extent to which information is clear and easily used
Understandability	extent to which data are clear without ambiguity and easily comprehended
Amount of data	extent to which the quantity or volume of available data is appropriate
Believability	extent to which information is regarded as true and credible
Navigation	extent to which data are easily found and linked to
Reputation	extent to which information is highly regarded in terms of source or content
Useful	extent to which information is applicable and helpful for the task at hand
Efficiency	extent to which data are able to quickly meet the information needs for the task at hand
Value added	extent to which information is beneficial, provides advantages from its use

5. METRICS OF DATA QUALITY

To be able to measure the quality of data, assessments along a number of dimensions is necessary [15]. Most of the DQ attributes depend on user assessment, which may be dependent on the user perception. However, other attributes are associated with the data itself at the table or record level. The assessment of some attributes is related to the metadata constraints level as shown in Figure 2 below. In the following sections, a description of the attributes in each category will be shown.

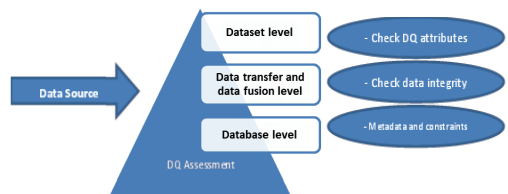


Figure 2. Data Quality Assessment process

5.1 Data and Metadata Related

Within a specific dimensional category, the specific measure to assess a specific dimension could vary from organization to another [15]. However, we present some metrics that developed by [19] and used for DQ assessment by [15].

The assessment of data related attributes are: accuracy, consistency, validity. While metadata related attributes could be: completeness, data domain, data type. Other attributes' assessment is related to the user evaluations. These attributes are usability, believability, reputation, relevancy and other attributes listed in Table 1.

Some attributes can be assessed by using mathematical formulas, such as, completeness. Complete data has been defined as data having all values recorded [20]. The completeness dimension can

be viewed from at least three perspectives: schema completeness, column completeness, and population completeness [15]. An incomplete value represents an unknown or missing value in the real world, or it represents a value yet to be entered into a database, a value of null is used to represent an incomplete data item [21]. Therefore, it can be calculated in the simple formula as:

$$\text{Completeness} = 1 - (\text{No. of incomplete items} / \text{total no. of items})$$

An item could be a file or a record. Accuracy denotes the extent to which data are correct and free-of-error [8]. The dimension of accuracy itself, however, can consist of one or more variables, only one of which is whether the data are correct [15]. If one is counting the number of data units in error, then the metric is:

$$\text{Accuracy} = 1 - (\text{no. of items in error} / \text{total no. of item})$$

5.2 User Related

As mentioned, most of DQ attributes could be assessed using user evaluation. A value expressed in the range of 0 to 1 has long been held as a desirable trait for DQ metrics [22]. Therefore, quantifiable measurements of specific DQ variables must be performed.

In our assessment, the user or consumer evaluation of the data should be considered. The assessment will be in the form of a set of questions that will be answered by the consumer. These attributes are: usability, accessibility, amount of data and other attributes that are evaluated by user rating. The user should select one value from a range of values for each attribute.

We need to mention here that the source of the data has a significant impact to DQ [23]. In general, there are two sources for data production: Manual (Human) and Automatic (Machine) data source. As studied before, most of issues regarding DQ are resulted from Manual data sources [23]. This attribute is represented by Reputation, and the user will be asked about the data source and a corresponding value will be given for each source of data.

6. Multi-Criteria DQ Assessment

6.1 Attributes Ranking

In order to choose the attributes and the appropriate weights that will be used in the DQ measuring process, Analytical Hierarchy Process (AHP) method will be used. The Analytic Hierarchy Process (AHP) has been developed by T. Saaty (1980). AHP is one of the most popular and powerful techniques for decision making in use today [24]. It allows the use of qualitative, as well as quantitative criteria in evaluation.

The AHP is about breaking a problem down and then aggregating the solutions of all the sub-problems into a conclusion. Therefore, the AHP is a multiple criteria decision-making tool that has been used in almost all the applications related with decision-making. It helps the analyst to organize the critical aspects of a problem into a hierarchical structure similar to a family tree [25].

The AHP is built on a human beings' intrinsic ability to structure his perceptions or his ideas hierarchically, compare pairs of similar things against a given criterion or a common property, and judge the intensity of the importance of one thing over the other [24]. AHP employs pairwise comparison in which experts

compare the importance of two factors on a relatively subjective scale. In this way a judgment matrix of importance is built according to the relative importance given by the experts.

AHP includes some steps as [26]:

1. Structuring the decision problem and selection of criteria: by arranging all the components in a hierarchy provides an overall view of the complex relationships and helps the decision maker to assess whether the elements in each level are of the same magnitude so that they can be compared accurately.
2. Priority setting of the criteria by pairwise comparison (weighing): For each pair of criteria, the decision maker is required to respond to a question such as "How important is criterion (A) relative to criterion (B)?" Rating the relative "priority" of the criteria is done by assigning a weight between 1 (equal importance) and 9 (extreme importance) to the more important criterion. The weighing is then normalized and averaged in order to obtain an average weight for each criterion
3. Pairwise comparison of options on each criterion (scoring): For each pairing within each criterion the better option is awarded a score, again, on a scale between 1 (equally good) and 9 (absolutely better), whilst the other option in the pairing is assigned a rating equal to the reciprocal of this value. Each score records for each option. Afterwards, the ratings are normalized and averaged.
4. Obtaining an overall relative score for each option: In a final step the option scores are combined with the criterion weights to produce an overall score for each option. The extent to which the options satisfy the criteria is weighed according to the relative importance of the criteria.

However, with AHP, each element in the hierarchy is considered to be independent of all the others. Many attributes of the quality of data are related. Therefore, an attempt to categorize the attributes to be in independent groups has been done.

6.2 Ranking Aggregation Method

In this study, the DQ metrics were defined. The AHP method described before is used to obtain an overall quality score for the provided data set.

The four steps of AHP were applied for weighing and scoring of the DQ attributes. The overall score of DQ will be found using the weighted summation. In this case the values of n single variable metrics will be aggregated using the following [15]:

$$DQ = \sum_{i=1}^n (a_i \cdot M_i) \quad (1)$$

where a_i is a weighting factor, $0 \leq a_i \leq 1$, and $a_1 + a_2 + \dots + a_n = 1$. M_i is a normalized value of the assessments of the i th attribute.

In order to apply AHP method to the attributes of the data described in section (3), the total DQ assessment hierarchy has been created as described in Fig 3 below. These attributes has been categorized to be more independent. After that, a questionnaire was conducted to collect the weighing and scoring values from different users using a sample dataset as a case study to validate the suggested method.

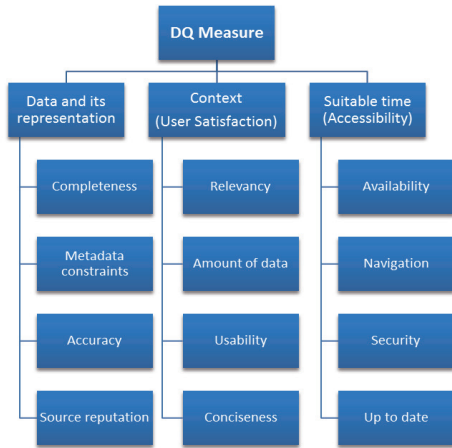


Figure 3. Overall DQ Assessment hierarchy.

7. SAMPLE DATA COLLECTION

In order to validate the suggested model, the model has been applied to Ofelia data base provided by Trafikverket (The Swedish Transportation Agency). As discussed before, the metrics of DQ has been calculated. Some of these measures have been calculated mathematically by using the formulas discussed before. The other metrics that are mainly depending on the user evaluation have been evaluated by questionnaire submitted to a group of researchers at a University in Sweden. The selected researchers are familiar to Ofelia database as they have conducted much research on this data. The questionnaire was divided into two parts. The first part was developed to be used for evaluation of the weight or the importance of each attribute. This part mainly used the comparison between each pair of DQ attributes. The user should select the importance of each attribute over the other.

The second part of the questionnaire was developed in order to compute an evaluation value for each attribute by every user depending on the user experience. The expert should choose a value for each attribute that will give an indication for the attribute evaluation of Ofelia database. After that, all these information were collected. Following the AHP steps, the weighing and scoring process has been done. Expert choice program also used in order to find the graphical representation for these attributes. Finally, equation (1) was used to find the final estimation of overall DQ measure for the data used in this study.

8. RESULTS AND DISCUSSIONS

Using the described questionnaire and applying AHP method, the ranking of the DQ attributes has been shown in Figure 4. From Figure 4, we can see the ranking of all attributes according to the user evaluation. We can see that completeness is the most important aspect for users. A pairwise comparison has been done between each class of the categories as shown in Figure 5. After ranking the attributes, the users' evaluations have been also collected in order to fulfill all the requirements of equation (1). These evaluations are shown in Table 2. Using these values, the final estimation of DQ can be calculated.

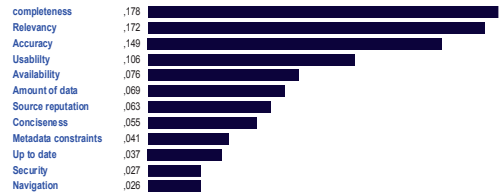


Figure 4. Attributes ranking.

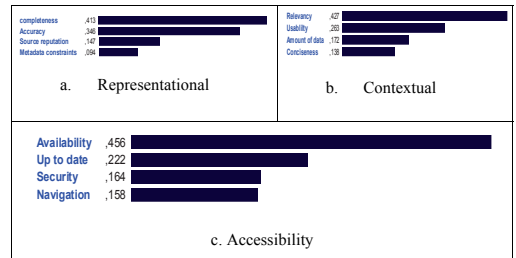


Figure 5. Pairwise comparison.

Table 2. Attributes user evaluation

Attribute	Evaluation
Completeness	7,5
Reputation	5
Relevancy	5,5
Accuracy	6,5
Up to date	8
Metadata Constraints	7,5
Availability	7,5
Usability	4
Amount of data	7,5
Conciseness	6
Security	7,5
Navigation	5

After applying equation (1), the overall score of DQ measure for the data of the study was (63.42%). Having that score, companies can know how the quality of their data is, what should be improved and what the most important attributes to the users are. In addition, that score will help the decision makers to know about the quality of their decisions made while using this data.

9. CONCLUSIONS

In order to improve DQ as well as to evaluate the current status, dimensions and attributes have to be measured. Multi-criteria DQ evaluation is a useful tool that may aid decision-making processes by providing a means by which information can easily assessed. As evidence and research have shown, information does not always reflect a high degree of quality or satisfy the intended need, which creates challenges during the utilization process and

delays in decision making. Furthermore, consequences of ineffective decisions and operational inefficiencies, are created as a result of poor quality information, affect an organization's performance.

DQ is composite of different dimensions which have different evaluation based on the data management and presence of trusted data sources. DQ measurement is based on subjective and objective measurement i.e. quantitative and qualitative measures. Qualitative and quantitative data evaluations are important to achieve reliable data collection. DQ evaluation starts with proper documentation, standards compliance and known metadata limits. In this study, the relevant attributes were identified where they describe if the right data are available with the appropriate event and with the needed context information. The proposed method for DQ measurement was applied. Although evaluating DQ poses many challenges, using the proposed evaluation methodology will provide decision makers and researchers important steps to identify and address DQ issues. However, this evaluation was mainly based on qualitative evaluation.

Thus, some challenges were addressed during this study. The problem with qualitative methods is that they are based on user evaluation, which may be affected by many factors such as expertise, context of usage, large difference from one user to another, etc. In the future research, more quantitative and statistical methods need to be used and the results can be compared with the qualitative ones. Moreover, both qualitative and quantitative methods may be merged in order to obtain more adequate evaluation. In the other hand, the method used in this study, AHP, suggests that the attributes should be independent. In this study, an attempt to make the attributes independent hierarchy was carried out. However, we need to use different ranking methods and compare the results to find the best method for attributes ranking and evaluations. The aggregation method also needs to be further investigated in order to make sure that it's the best choice.

10. REFERENCES

- [1] A.F. Karr, A.P. Sanil and D.L. Banks, "Data quality: A statistical perspective," *Statistical Methodology*, vol. 3, no. 2, pp. 137-173 2006.
- [2] T. Redman, "Data: An unfolding quality disaster," *DM REVIEW*, vol. 14, no. 8, pp. 21-23 2004.
- [3] W.W. Eckerson, "Data quality and the bottom line: Achieving business success through a commitment to high quality data," *The Data Warehousing Institute*, pp. 1-36 2002.
- [4] R. Marsh, "Drowning in dirty data? it's time to sink or swim: A four-stage methodology for total data quality management," *The Journal of Database Marketing & Customer Strategy Management*, vol. 12, no. 2, pp. 105-112 2005.
- [5] F. Naumann and C. Rolker, *Assessment methods for information quality criteria*, Citeseer, 2000.
- [6] M. Bovee, R.P. Srivastava and B. Mak, "A conceptual framework and belief-function approach to assessing overall information quality," *Int J Intell Syst*, vol. 18, no. 1, pp. 51-74 2003.
- [7] Y. Wand and R.Y. Wang, "Anchoring data quality dimensions in ontological foundations," *Commun ACM*, vol. 39, no. 11, pp. 86-95 1996.
- [8] R.Y. Wang and D.M. Strong, "Beyond accuracy: What data quality means to data consumers," *J.Manage.Inf.Syst.*, pp. 5-33 1996.
- [9] L.P. English, *Improving data warehouse and business information quality: methods for reducing costs and increasing profits*, Wiley New York, 1999.
- [10] Y.W. Lee, D.M. Strong, B.K. Kahn and R.Y. Wang, "AIMQ: A methodology for information quality assessment," *Information & management*, vol. 40, no. 2, pp. 133-146 2002.
- [11] R.Y. Wang, "A product perspective on total data quality management," *Commun ACM*, vol. 41, no. 2, pp. 58-65 1998.
- [12] M. Alkhattabi, D. Neagu and A. Cullen, "Assessing information quality of e-learning systems: A web mining approach," *Comput.Hum.Behav.*, vol. 27, no. 2, pp. 862-873 2011.
- [13] H. Zhu and H. Wu, "Quality of data standards: Framework and illustration using XBRL taxonomy and instances," *Electronic Markets*, vol. 21, no. 2, pp. 129-139 2011.
- [14] R.Y. Wang, "A product perspective on total data quality management," *Commun ACM*, vol. 41, no. 2, pp. 58 1998.
- [15] Y.W. Lee, L.L. Pipino, J.D. Funk and R.Y. Wang, *Journey to data quality*, MIT press Cambridge, Massachusetts; 2006.
- [16] K. Atkinson, "Accounting for data: A shortcoming in accounting for intangible assets," *Academy of Accounting and Financial Studies Journal*, vol. 10, no. 2, pp. 85 2006.
- [17] R.S. Taylor and M.J. Voigt, *Value added processes in information systems*, Greenwood Publishing Group Inc., 1986.
- [18] S. Knight and J. Bum, "Developing a framework for assessing information quality on the world wide web," *Informing Science: International Journal of an Emerging Transdiscipline*, vol. 8, no. 5, pp. 159-172 2005.
- [19] E.F. Codd, "A relational model of data for large shared data banks," *Commun ACM*, vol. 13, no. 6, pp. 377-387 1970.
- [20] P. Gomes, J. Farinha and M.J. Trigueiros, "A data quality metamodel extension to CWM," in *Proceedings of the fourth Asia-Pacific conference on Conceptual modelling-Volume 67*, 2007, pp. 17-26.
- [21] 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.
- [22] R.A. Dillard, "Using data quality measures in decision-making algorithms," *IEEE Expert*, vol. 7, no. 6, pp. 63-72 1992.
- [23] 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 : MPM²*, Luleå, Sweden, 2011, pp. 69.
- [24] E. Forman and K. Peniwati, "Aggregating individual judgments and priorities with the analytic hierarchy process," *Eur.J.Oper.Res.*, vol. 108, no. 1, pp. 165-169 1998.
- [25] T.L. Saaty, "How to make a decision: The analytic hierarchy process," *Eur.J.Oper.Res.*, vol. 48, no. 1, pp. 9-26 1990.
- [26] T.L. Saaty, *What is the analytic hierarchy process?*, Springer, 1988.

Big Data Mining in eMaintenance: An Overview

Liangwei Zhang
Division of Operation and
Maintenance Engineering,
Luleå University of Technology,
97187 Luleå, Sweden
+46 (0)920 49 13 82
liangwei.zhang@ltu.se

Ramin Karim
Division of Operation and
Maintenance Engineering,
Luleå University of Technology,
97187 Luleå, Sweden
+46 (0)920 49 23 44
ramin.karim@ltu.se

ABSTRACT

Maintenance related data are tending to be increasingly huge in volume, rapid in velocity and vast in variety. Data with these characteristics bring new challenges with respect to data analysis and data mining, which requires new approaches and technologies. In industry, related research and applications, some contributions have been provided to utilize Big Data technologies for extraction of information through pattern recognition mechanisms via eMaintenance solutions. Today, the existing contributions are not enabling a holistic approach for maintenance data analysis and therefore are insufficient. However, the immense value hidden inside the Big Data in eMaintenance is arousing more and more attention from both academia and industry. Hence, this paper aims to explore eMaintenance solutions for maintenance decision-making through utilization of Big Data technologies and approaches. The paper discusses Big Data mining in eMaintenance through a general manner by employing one of the widely accepted frameworks with the name of Cross Industry Standard Process for Data Mining (CRISP-DM). In addition, the paper outlines features of maintenance data and investigates six sub-processes (i.e. business understanding, data understanding, data preparation, modeling, evaluation and deployment) of data mining applications defined by CRISP-DM within the domain of eMaintenance.

Keywords

Data Mining, eMaintenance, Big Data, CRISP-DM

1. INTRODUCTION

Nowadays, with the rise of numerous types of sensors, mobile devices, tether-free, web-based applications and other information and communication technologies (ICT), the ability of asset-intensive enterprises to procure and store maintenance data has been enhanced. The large troves of maintenance data can be decentralized, fast-flowing, heterogeneous and complicated to handle which reflects features of the emerging term Big Data [1], which is defined by Gartner Group [2] as: "high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization."

Maintenance data with Big Data characteristics pose new challenges to conventional data mining process which is one of the main research topics in eMaintenance domain. Kajko-Mattsson et al. [3] referred eMaintenance as: "a multidisciplinary domain based on maintenance and information and communication technologies ensuring that the eMaintenance services are aligned with the needs and business objectives of both

customers and suppliers during the whole product lifecycle". Mining this securable data is troublesome in virtue of that most of these data are semi-structured or unstructured. Traditional computing techniques like relational databases, designed to cope with structured data, do not fit well to the new circumstances [4]. Furthermore, real-time stream data reflecting the status of equipment need to be continuously monitored [5]. Timely analysis and response should be assured when disturbance occurs, especially to the equipment of the utmost concern.

Data mining is capable of supporting decision-making, and this is being researched upon in the context of eMaintenance for creating competitive advantage for enterprises and organizations [6], [7], [8]. The use of data mining techniques in maintenance started around the 1990s [9], [10], subsequently lots of efforts have been made by researchers on this topic. Typical applications include maintenance strategy optimization, fault diagnosis and failure prognosis. Unfortunately, owing to the limitation of capacity in data storage and processing in the past, many works were focusing on mining sampling data which may result in losing details [11]. With the development of non-relational database, distributed file storage system, cluster computing and other Big Data technologies, it is possible and imperative to conduct data mining to the huge maintenance datasets from the perspective of Big Data. In the context of eMaintenance, a few cases applying Big Data technologies to support maintenance decision-making can be found. Such as Bahga et al. [12] applied Hadoop Distributed File System (HDFS) and MapReduce parallel programming model to form a large-scale distributed batch processing infrastructure to assist fault prediction. Stream data processing technique combined with Telematics were harnessed to analyse real-time data collected from sensor arrays then enable condition-based maintenance [13]. By reviewing existing literatures, this paper intends to explore eMaintenance solutions for maintenance decision-making through utilization of Big Data technologies and other auxiliary approaches.

Quite a few standard processes have been developed to guide the implementation of data mining projects. CRISP-DM is one of the popular methodologies with befittingly systematized, structured and defined procedures, allowing that a project could be easily understood and revised [14]. A project guided by CRISP-DM model commonly consists of six phases [15], including business understanding, data understanding, data preparation, modeling, evaluation and deployment. In this paper, we will discuss Big Data mining in eMaintenance within the frame of CRISP-DM.

The remainder of this paper is structured as follows. In section 2, we present the constitution and features of Maintenance data then align it to Big Data characteristics. Section 3 reviews existing literatures within the frame of CRISP-DM process and explores

potential eMaintenance solutions through utilizing Big Data technologies. As depicted in section 4, current challenges of big data mining in eMaintenance are formidable and more studies are expected. Finally, we review and conclude the paper in Section 5.

2. Characteristics of “big maintenance data”

In accordance with the European standard (EN 13306:2001) [16], maintenance is a combination of all technical, administrative and managerial actions during the life cycle of an item, intended to retain it or restore it to, a state in which it can perform the required functions. E-maintenance, on a higher abstraction level, was also defined as the maintenance managed and performed via computing [3]. When developing eMaintenance solutions, the prevalently concerned analysis object of the aforementioned computing is maintenance data, and this entails a sufficient dissection to maintenance data.

Traditionally, maintenance data are regarded as a by-product of this process and can be leveraged in return to augment efficiency of maintenance activities. In order to discuss data mining from maintenance datasets in the context of Big Data, we elaborate maintenance data and its' constitution in the first place then align features of maintenance data to Big Data characteristics in this section.

2.1 Maintenance data anatomy

Maintenance data have a wide coverage and a great variety of forms. It can be as specific as a maintenance work order, or as generic as the maintenance strategy and objectives. Examples of maintenance data include asset ledgers, drawings, contracts, licences, legal agreements, regulatory and statutory requirements, standards, safety and hazard documents, technical instructions, procedures, operating criteria, asset performance and condition data, or all asset management records. In the following, the composition and features of maintenance data will be discussed from different angles.

2.1.1 Maintenance data composition

Maintenance data don't merely consist of the data generated during the course of carrying out maintenance tasks but also constitute all the relevant data produced before and after maintenance. From the perspective of asset life cycle management, maintenance data can be broadly divided into four categories corresponding to the phases of asset life cycle [17] during which it has been generated.

- Stage 1: creation, acquisition or enhancement of assets;

Usually, people would think that maintenance data have no direct linkage with this phase, because events occurred in this period are less relevant with maintenance tasks to some sense. However, the data like technical specifications provided by the asset supplier can exert a great impact on the maintenance strategy. For example, the recommended preventive maintenance interval should be considered as a significant input to determine the frequency of implementing the periodical maintenance tasks. More maintenance data produced in this phase include asset drawings, technical specification documentations, number of attached spare parts, purchase contracts and warranty terms, etc.

- Stage 2: utilization of assets;

Asset utilization schedule may affect the implementation of maintenance work, and vice versa. For example, some maintenance tasks are executed by the opportunity when partial or full stoppage of a process area occurs provoked by asset failure or scheduled shutdown. Data produced in this phase that could be regarded as maintenance data constitute production scheduling, hazard and precaution identifications, environmental conditions, operating load and parameters, etc.

- Stage 3: maintenance of assets;

The great mass of maintenance data are generated in this phase and it can be leveraged in return to continuously support maintenance work. In this phase, maintenance data encompass condition monitoring data, feedback of routing inspection, failure data, maintenance resources, work order, overhaul and refurbish plan, and so forth.

- Stage 4: decommissioning and/or disposal of assets

Decommissioning or disposal of assets is mainly associated with financial accounting in terms of assets depreciation or scrapping. Maintenance data in this phase principally contain depreciation expense that should be taken into account when calculating the Life Cycle Cost (LCC) of an asset. The information of recycling a retiring asset also needs to be thoroughly recorded in case of further investigation.

2.1.2 Multiple sources of maintenance data

Sources of maintenance data are numerous. It could be the historical maintenance records in a Management Information System (MIS), or troubleshooting guidelines on a printed user manual, or even public maintenance standards derived from webpages on the Internet.

Although quite a few maintenance data are still non-digital, which depends on the informatization degree of the concerned company, we shall anticipate a tendency of digitising these non-digital data. In this section, we will discuss some of the prevailing systems where these digital maintenance data originates.

- Enterprise Asset Management (EAM)

EAM, originally known as Computerized Maintenance Management System (CMMS), aims to fulfil life cycle management of the physical assets of an organization. It is designed to drive indirect value chain by tracing all events occurred during assets' whole life span, which covers design, construction, commissioning, operation, maintenance and decommissioning/disposal.

- Enterprise Resource Planning (ERP)

ERP is a suite of management software that can be used for a company to store and manage data from every stage of business. ERP software drives direct value chain by tracking business resources (cash, raw materials, labours, etc.) and the status of business commitments.

With the growing demand on asset management, many ERP software vendors have integrated asset management module (SAP PM Module, Oracle EAM module and so forth) into existing packages.

- Condition monitoring System (CMS)

CMS intends to keep continuous or intermittent surveillance on critical components or units then enable predictive maintenance. The collected parameters of condition in machinery (velocity, vibration, temperature, etc.) are then analysed to identify potential faults that could result in major failures.

- Supervisory Control and Data Acquisition Systems(SCADA-Systems)

SCADA is a system for collecting real-time data, controlling processes and monitoring equipment from remote locations [18].

- Safety Instrumented System (SIS)

SIS is a distinct and reliable system to implement one or more safety instrumented functions, which must have achieved a specific Safety Integrity Level (SIL), for protecting personnel, general public or environment [19].

As stated above, maintenance data can be distributed in multiple information repositories. Owing to the lack of long-term planning in organizations, only temporal demands were taken into account at different stages of information platform construction. Afterwards, these standalone systems form the so-called information islands where large amount of redundant data and junk information exist. The inconsistency of data dispersed on these systems leads to the difficulty in exchanging data. In such circumstance, several integrated eMaintenance platforms were developed to either fuse data into a centralized data warehouse or facilitate data exchange across these individual systems. The PROTEUS project developed a platform, which enables data exchange and interoperability of the diverse systems. The platform was designed to support remote diagnosis and maintenance [20]. Another example is the Computer Aided Safety and Industrial Productivity (CASIP) platform which dedicated to condition monitoring and remote diagnosis [21].

2.1.3 Various formats of maintenance data

As with most other application fields, maintenance data could be represented by multitudinous formats. In general, it can be categorized into non-digital data and digitized data. Non-digital data may be manuscript written on a paper, or video recorded on a tape, or even maintenance experience and knowledge that exist only in the mind of a skilled maintenance engineer. With respect to digitized data, it can be classified into structured, semi-structured and unstructured data.

- Structured data

Structured data are consistently preserved in a well-defined schema and it conforms to some common standards. These

standards restrict what fields of data can be stored and how it will be stored, such as data type, data size and constraints.

Structured data are more likely to be stored in relational databases (Oracle, DB2, SQL Server, etc.) and could be easily accessed and manipulated by Structured Query Language (SQL). Most of maintenance data curated in above-mentioned information systems are structured.

- Unstructured data

Unstructured data refer to data without a predefined schema. It cannot be smoothly classified and fitted into conventional relational databases thus hard to be processed [22]. It is estimated that almost 95% of all digital data are unstructured or semi-structured [11].

Here are some examples of unstructured data in eMaintenance domain: technical documentations of equipment, images taken by infrared thermal imager, frequency spectrum collected by vibration detection equipment, movement videos captured by high-speed digital cameras, etc.

- Semi-structured data

Semi-structured data are a cross between above two types of data. It is primarily text based and conforms to some specified rules. Within semi-structured data, tags or other forms of markers are used to identify certain elements. In comparison with structured data, semi-structured data have a relatively loosened structure.

Extensible Markup Language (XML) document is a typical format of semi-structured data file. XML has played a great role in encoding and sharing data between different applications and it has also been widely used in manufacturing industry [23]. With a more flexible data model, XML file cannot be processed in the manner of how data are manipulated in relational databases. Accordingly, NoSQL databases were put forward to manage these semi-structured data and proven to be more scalable and agile [24]. More details about NoSQL databases will be discussed in later section.

2.2 Aligning maintenance data features to Big Data characteristics

There is not a consensus in the definition of Big Data so far, one of the most cited definitions of Big Data was given by the Gartner Group [2] (see section 1). It also described the characteristics and challenges of Big Data by using the three “Vs”, which are high volume, high velocity and high variety. Based upon the elaboration to maintenance data in section 2.1, let us embark on aligning the features of maintenance data to Big Data characteristics.

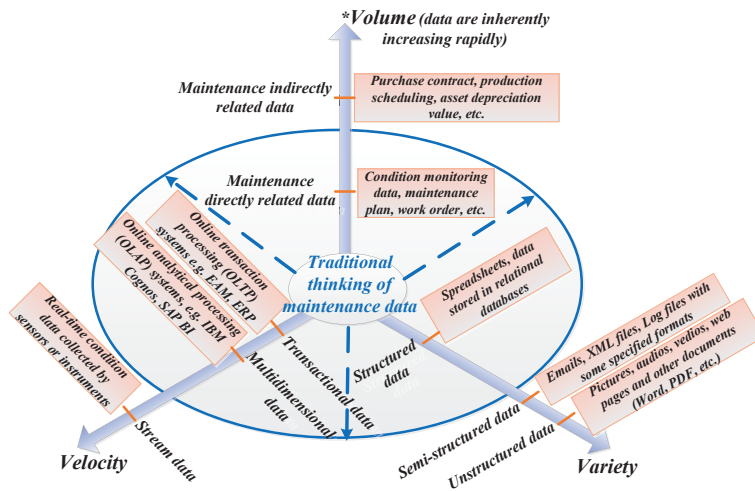


Figure 1. Characteristics of maintenance data

- Volume

The scale of data is tremendously increased in the past decade [25]. It is not only the most notable feature of Big Data but also the nutriment that facilitates the development of information technology which enable us to extract value from data. Without exception, maintenance data are now undergoing rapid growth as well. Moreover, the range of maintenance data should be expanded to a wider space containing all data that have been generated in the interrelated business process of maintenance (See the volume dimension in Figure 1).

The growing volume of maintenance data is demonstrated with an example below: In a Swedish hydropower plant, the real-time condition data (vibration, temperature, pressure, etc.) of one generator are measured by transducer arrays deployed on it. At each second, condition data are captured and transmitted through 128 analog channels and 64 digital channels then saved in a data logger database. More than 30 million tuples are preserved in one year for this single generator. It is not hard to imagine how big the data are if we consider all the equipment in the plant and the data from other sources. This example also illustrates the next characteristic of Big Data: rapid velocity.

- Velocity

The rate at which maintenance data are flowing into the line of business has placed higher requirements to the company. Prompt analysis to real-time data will facilitate decision-making in a quick response manner that may allow organizations seize opportunities in the dynamic and changing market, sometimes avoid too much loss.

As shown in the velocity dimension of figure 1, originally the requirements to the speed of maintenance data flow contain two aspects. Firstly, transactional data ensure basic functionality and performance of Online Transactional Processing (OLTP) systems under concurrency without compromising the principle of relational databases: ACID (Atomic, Consistency, Isolation and

Durability). In the second place, multi-dimensional data are created to allow complicated analytical and ad hoc queries within a tolerable execution time. The velocity of an OLAP system is speeded up by introducing data redundancy in data cube, which means size of data will be augmented immensely.

Large troves of real-time maintenance data stream are originating from different types of sensors. Equipment anomalies indicated by those stream data should be addressed promptly and agile actions are required to avoid unplanned breakdown and economical loss. Here velocity indicates not only the data are fast moving, but also the intelligent diagnosis and subsequent human intervention should be swift.

- Variety

It points to the types of maintenance data owned by organizations are becoming increasingly diverse. Maintenance data could be derived from wireless sensor network, running log documents, surveillance image files, audio and video clips, complicated simulation and GPS enabled spatiotemporal data, and much more. Most of these maintenance datasets are thoroughly unstructured or semi-structured and hence difficult to curate, categorize and analyse using traditional computing solutions.

In addition, another characteristic of Big Data commonly embraced by researchers is the fourth dimension "Veracity" [26], referring to the messiness, trustfulness of data. In the context of maintenance data, it can be inaccurate (e.g. sensor data with environmental noise), uncertain (e.g. prediction of the remaining useful life of a critical component) and biased (e.g. interval of time-based maintenance). This issue is normally addressed as the data quality problem and tackled by data quality assurance/control procedures and data cleaning techniques.

After revealing the features of the computing target (i.e. maintenance data) in eMaintenance solutions, we are ready to probe into how to extract patterns from these complex data

repositories and explore eMaintenance solutions by applying Big Data technologies and other auxiliary approaches.

3. Mining the “big maintenance data”

Many contributions have been made to formulate a standard process that is able to guide the implementation of data mining applications. Knowledge Discovery in Databases (KDD) process [27], in which data mining is considered as one of the phases, was put forward in 1996. SEMMA process [14] developed by the SAS Institute was integrated into one of SAS core modules Enterprise Miner. CRISP-DM is another popular methodology proposed by a consortium composed with several European companies including DaimlerChrysler, SPSS, NCR, OHRA, etc. CRISP-DM provides detailed neutral guidelines for data mining implementation that can be easily duplicated and revised [28]. It has been validated to be a very helpful process through many real-world applications [29], [30].

In this study, CRISP-DM is employed as the methodology to guide the process of Big Data mining in eMaintenance domain. The lifecycle of the CRISP-DM comprising six phases: business understanding, data understanding, data preparation, modeling, evaluation and deployment. Figure 2 shows above six phases and lists some of the tools and techniques that can be utilized in each phase of the CRISP-DM lifecycle. The following sections are organized correspondingly to the six phases.

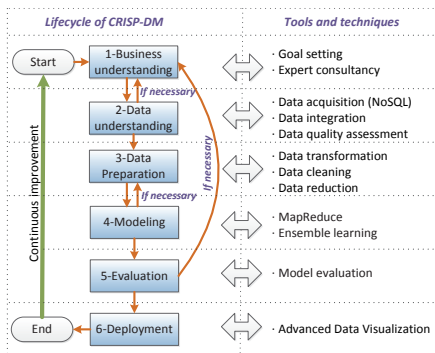


Figure 2. Lifecycle of the CRISP-DM methodology

3.1 Phase1: Business understanding

This initial phase focuses on understanding the project objectives and requirements from a business perspective, then converting this know-how into a data mining problem definition and a preliminary plan designed to achieve the objectives [15].

In this phase, data mining goals are stated based on the business objectives. Ignoring this step is to expend a great deal of efforts on producing the right answers to the wrong questions. For example, after some complex modelling and computing one might produce the reorder point and economic order quantity for a particular type of spare part but neglecting that the spare part is sold on consignment and no extra cost is caused except for management fee.

Domain experts’ experiences and knowledge is introduced at this stage and it usually has a significant impact on the applicability of the final results. Such application domains’ knowledge can also provide additional information to guide Big Data algorithm designs as well as add technical barriers to the whole process [1]. As suggested by the CRISP-DM guidance, more detailed fact-finding about resources, constraints, assumptions and other aspects that could possibly affect subsequent data mining tasks should be assessed as early as possible. In the context of the eMaintenance field, fault diagnosis and failure prognosis is a typical application in data mining projects. By taking this as an example, we will elaborate what normally should be considered to the application in this phase.

Fault diagnosis and failure prognosis is a process to detect product degradation based on on-line sensing data then predict potential failures of monitored components. It enables condition-based maintenance (CBM) and constitutes the hard-core of an eMaintenance system [31], which is referred to a predictive maintenance system that provides only monitoring and predictive prognostic functions. Before analyzing the data, it is imperative to preview the business concerns including but not limited to the items listed below.

- Common issues such as maintenance objectives, system decomposition, asset criticality, fault tree, environmental conditions, equipment installation process, existing maintenance strategy, etc.
- On-site system health monitoring including sensor types and models, sampling rate, data infrastructure, network configuration, etc.
- Logistics supports including priority of maintenance work, alarm lead-time, maintenance labour availability, spare part availability, maintenance tool, outsourcing service provider, etc.

Normally, a project plan describing how to achieve the data mining goals is formed at the end of this phase. Since it is beyond the scope of this paper, it will not be covered consequently.

3.2 Phase 2: Data understanding

The data understanding phase begins with an initial data collection and proceeds with describing data, exploring data and verifying data quality problems [15]. In the long run, comprehensive investigation to the data characterization is critical to the whole data mining process, for example, how much the maintenance data are going to be and how fast the maintenance data are likely to grow and change will fundamentally determine the choice on data infrastructure and algorithm design.

Considering the recommended tasks in this phase, the following contents cover three frequently used eMaintenance tools: data acquisition, data integration and data quality assessment.

3.2.1 Data acquisition

In correspondence to the task “initial data collection”, data acquisition is a key step that enables maintenance decision support in eMaintenance solutions. Data infrastructure is one of the significant factors that could affect the efficiency of a data acquisition system. Whereas data infrastructure of the days before Big Data is poorly fitted to the requirements of staggering amount of maintenance data, which has an increasing rate of change. For

example, the newly deployed sensors might require an alteration to the table structure of the Relational Database Management System (RDBMS) where the previous data are stored. However, any modifications and reconfigurations to the predefined schema in a RDBMS will lead to downtime and possibly extra labour cost. Non-relational NoSQL model is a Big Data technology that can deal with the huge, fast-moving and frequently changed maintenance data.

NoSQL databases were developed in response to a rise in the huge volume of data, the frequency in which the data are accessed, and high-performance processing needs. In contrast with traditional RDBMS, NoSQL database provides high throughput, horizontal scalability, replication and partition capability, dynamic schema adaptability, while retaining eventual consistency [32].

NoSQL databases can be generally classified into four categories: key-value stores, document stores, column family stores and graph stores [33]. Each of these NoSQL or relational databases has its' strength and shortages thus a "one size fits all" database is unavailable to all applications. Organizations must select the most suited database to underpin its' own applications based upon the requirements like consistency, availability, partition tolerance, etc.

Several cases using NoSQL databases have been reported in eMaintenance domain. Such as Thantriwatté et al. [34] used Redis database (a key-value store) to construct a query processing system for wireless sensor networks. Traditional data storage solution was substituted with Cassandra database (a column family store) in the ATLAS PanSA monitoring system [35].

3.2.2 Data integration

Maintenance data are acquired from multiple data sources (See section 2.1.2) and the value of data explodes when data are linked with each other. Therefore data integration is a crucial step in creating value from maintenance data.

In figure 3, we contrasted the difference of data integration solutions between Big Data mining and traditional data mining. Traditional solution of data integration would firstly build a centralized data warehouse. Then data from the heterogeneous sources are extracted, transformed and loaded (ETL) into the specified data warehouse with a compatible schema [36]. For instance, P. Bastos et al. [37] developed a predictive maintenance system to forecast potential failure based on the data dispersed in different facilities. In the era of Big Data, with the development of Distributed Storage System (DSS) and Cluster Computing, it is no longer advisable to move such large quantity of maintenance data considering the required I/O time and cost, in other word, moving computation to data is cheaper than the way around. Distributed storage systems provide scalable data storage with inexpensive commodity servers. Cluster Computing provides functionality like parallel computing, work load balancing and fault tolerance [38].

A commonly used Big Data technology, Hadoop framework, suggests that large datasets can be organized and processed while retaining the data on the initial data storage repositories. Hadoop framework consists of two core components: HDFS and MapReduce. HDFS is a storage system aims to store very large datasets in a distributed manner, and then provide high-performance access to the data in Hadoop clusters [39]. Hadoop MapReduce is a programming paradigm for processing massive data across multiple servers in a Hadoop cluster [40]. Hereby, data

integration of Big Data mining is essentially incorporated into the infrastructure of data storing and processing paradigm.

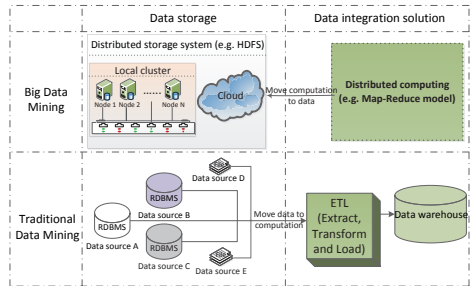


Figure 3. Comparison to data integration solutions

3.2.3 Data quality assessment

The quality of data will tremendously influence the outcome of data mining applications because of the principle "Garbage in, Garbage out (GIGO)" [41]. Numerous data mining algorithms require quality data, such as outliers may dramatically distort the output of Pearson Product-moment Correlation method which is sensitive to extreme values [42]. Many researchers have dedicated to the study of data quality assessment and several methodology and techniques have been proposed and applied into practice. For example, Batini et al. [43] described and compared thirteen different methodologies of data quality assessment and further identified open problems in data quality.

Maintenance data quality, for the reason of its' intrinsic complexity, has not been discussed adequately in the prior research [44]. Maintenance data in various information systems can be incomplete, noisy and inconsistent, due to many reasons. For example, the concepts of personnel who built up the system would determine the definition of the hierarchical asset structure, which can differ greatly between an asset management point of view and a financial management perspective. Besides, other factors (e.g. measurement error, typing mistake, inappropriate work flow design, etc.) may also lead to bad data quality.

The scope of maintenance data has been broadened from the perspective of Big Data. Emails, webpages, log files, images, videos and other semi-structured or unstructured data may be even more chaotic due to the loosened structure. If an organization has substituted RDBMS with any NoSQL databases, the schema-free design might lead to more impure data unless inputting data are strictly verified and regulated in the application programs. For example, the dynamic schema in MongoDB (a document NoSQL database) allows documents (rows) in the same collection (table) have different set of fields (columns) or structure, and common fields (columns) in a collection's (table) documents (rows) may hold different types of data [45]. With the rising attention to unstructured and semi-structured data in the context of Big Data mining, assessment techniques to data quality are becoming increasingly complex [43]. It is worthwhile to conduct a comprehensive assessment to the data quality problem then suggest how to improve the data governance policy.

3.3 Phase 3: Data preparation

The data preparation phase covers all activities to construct the final dataset from the initial raw data [15]. In general, it is a very time-consuming step, usually claiming about 75% of the total data mining process [46]. Based upon the problem definition and determined data sources in the preceding steps, specific subsets of data are selected in this phase. Afterwards, several techniques (such as data transformation, data cleaning and data reduction, etc.) are applied to analyse raw data so as to yield quality data. The refined high quality data assure the accuracy of the final results. In corresponding to the tasks of this phase, frequently used techniques including data transformation, data cleaning and data reduction are discussed as below.

3.3.1 Data cleaning

In section 3.2, we have justified that maintenance data may be incomplete, noisy and inconsistent. Therefore data should be purified before feeding to the succeeding algorithms. Because of the necessity of expert involvement in rectifying inconsistent data, we will only present an overview to recent studies on handling missing values and outliers.

Missing values can be frequently seen in maintenance data, either from the system or man-made. Multiple methods and software have been developed to cope with missing values [47]. Data imputation is one of the common approaches dealing with missing values without discarding possibly useful information. Usually, statistical techniques and machine learning algorithms lay the foundation for imputing absent data. In [48], different methods of data imputation were compared and the result showed that the selected machine learning algorithms have significantly outperformed the specified statistical techniques, yielding more accurate prediction.

Outlier detection and processing is another typical approach to clean data. Hellerstein [49] summarized various techniques in cope with different types of outliers which includes univariate outliers, multivariate outliers, time series outliers and frequency outliers. Big Data tend to be sparse in high dimensional data so the notion of proximity loses its meaningfulness [50]. More robust distance functions need to be developed to detect high dimensional outliers. Subspace outlier detection approach was proposed to tackle outliers of high dimensional data and proven to be superior than traditional full-dimensional approach [51]. Another example is that an angle-based model dealing with high dimensional data was developed to measure variance in the angles between different vectors of points [52].

Maintenance data are multitudinous in formats. It is necessary to preprocess semi-structured and unstructured maintenance data before feeding them to models as well. In contrast with structured data cleaning, semi-structured data are troublesome to be dealt with by computers due to the complexity of human language. Fortunately, Natural Language Processing (NLP) has attracted lots of attention and been extensively studied during the past decade, several techniques like normalization, stemming and lemmatization, feature extraction, etc. are available to preprocess textual data [53]. Although insufficient, we can now find that some research — such as deep learning, aims to detect features from unstructured data — was reported recently [54], [55].

3.3.2 Data transformation

Data transformation is the process of converting or consolidating raw data into required form that is appropriate for mining. Commonly used methods in data transformation comprise: smoothing, aggregation, generalization, normalization and attribute construction [56]. To facilitate data transformation in the context of Big Data, Kandel et al. [57] developed an interactively visual tool Wrangler, which blends a mixed-initiative user interface with a declarative language. Prekopcsak et al. [58] conducted data transformation on a data warehouse architecture (Apache Hive) built on top of Hadoop by using a SQL-like query language HiveQL. Maintenance data are messy, and thereby need to be transformed to a neat and orderly form.

3.3.3 Data reduction

Data reduction refers to techniques that can be applied to simplify dataset into a smaller representation without losing significant information [56]. Generally, mining on a reduced dataset should be more efficient yet result in the same or approximate analytical results. Five strategies for data reduction have been identified by Han et al. [56] including data cube aggregation, attribute subset selection, dimensionality reduction, numerosity reduction, discretization and concept hierarchy generation.

In the eMaintenance domain, real-time sensor data are procured to exhibit asset health status and enable condition monitoring. High dimensional sensor data exert increasing pressure not only on data transmission and storage, but also on data analyzing because of the “curse of dimensionality”. Several studies have been carried out concerning data reduction on sensor data. The Least Mean Square algorithm was adopted to reduce the amount of sensor data to be transmitted while assuring the data could be reconstructed at certain accuracy [59]. A wavelet-based feature selection method was proposed for efficient data reduction to wireless sensor network in a structural health monitoring system [60]. An interval censoring mechanism together with a quantization method was applied to reduce amount of sensor data and deal with uncensored measurements [61].

3.4 Phase 4: Modeling

In this phase, various modelling techniques are selected and applied and their parameters are calibrated to optimal values [15].

Normally, several techniques can be applied to the same data mining problem types. For example, both Neural Networks (NN) and Support Vector Machines (SVM) are able to identify which of a set of classifications a new observation belongs to. Accordingly, applicable modelling techniques need to be assessed based on some predefined criteria, like interpretability, predictive accuracy and computational efficiency, which have been used as the criteria to evaluate different models [62]. Since plentiful efforts have been made on comparing various algorithms [63], [64], [65], pros and cons of each individual algorithm will not be covered in this paper. In this section, we will mainly discuss the fact that Big Data technologies can be harnessed to boost the execution efficiency of data mining algorithms.

Owing to the limitation of a single computer system’s physical memory space, execution of an algorithm to an extremely large dataset may be quite time-consuming, sometimes even impossible. For instance, the learning time of the rule-based classification algorithm C4.5 will grow rapidly with the expansion of the

training dataset size [66]. Hereby researchers have attempted to divide large dataset into subsets with a relatively small size for a single processor to handle then feed the subsets to algorithms for computing in parallel [67]. However, the architectures that enable the parallel execution in the past are expensive and old-fashioned parallel algorithms are usually lack of scalability and capacity of balancing load.

With the advent of cloud computing environment and Big Data technologies, the high cost of parallel execution infrastructures which was once one of the barriers to broaden application of parallel execution has now been reduced [68]. As mentioned in section 3.2, MapReduce is a distributed programming paradigm proposed by Google for parallel data processing on large shared-nothing clusters. It consists of two steps: Map and Reduce. Map step retrieves data from the distributed file system (e.g. HDFS), and then distribute the data as a set of key-value pairs to several computing nodes for processing. Once Map tasks are finished, the Reduce step begins to aggregate results with the same key that are generated during the Map step [40].

There are some researchers dedicated to parallelizing and speeding up the execution of data mining algorithms. In 2006, Chu et al. [69] employed MapReduce to parallelize a series of learning algorithms including locally weighted linear regression, k-means, logistic regression, naive Bayes, SVM, ICA, PCA, Gaussian discriminant analysis, EM, and back propagation. In 2009, Wu et al. [70] proposed a new method MReC4.5 in which MapReduce and C4.5 algorithm were synthesized to enhance effectiveness of the constructed decision tree. In 2012, Geronimo et al. [71] developed a parallel genetic algorithm using MapReduce to handle massive requirements to computational resources in intensive search-based technique. In 2014, He et al. [72] introduced MapReduce to the spatial clustering algorithm DBSCAN which was modified to become more scalable and efficient.

Computational efficiency is a significant factor to evaluate several applicable algorithms in a specific data mining application. By reviewing the existing research, we conclude that Big Data technologies can be applied to enhance computing efficiency of algorithms. Nevertheless, it is worthwhile to note that once Big Data technologies are utilized to parallelize tasks of algorithms, other potential criteria (such as scalability, load balance across computing nodes) must also be taken into consideration when evaluating a modelling technique. Additionally, to enhance accuracy of the data mining results, ensemble learning that utilizes multiple models to reduce the variance has received extensive concerns and studies. It has been proved to be more effective and accurate compared with single model learning [73].

Although lots of researchers are devoted to improving the performance of data mining techniques, applications in the eMaintenance domain that utilize the big data technologies to improve algorithms' execution efficiency and accuracy are scarce. To date, only one report can be found in eMaintenance field harnessing Big Data technology to parallelize tasks of data mining algorithms, i.e., Arshdeep et al. [12] adopted MapReduce to analyse machine sensor data to create and update case-base for fault detection.

3.5 Phase 5: Evaluation

At this stage the model (or models) obtained are more thoroughly evaluated and the steps executed to construct the model are reviewed to be certain it properly achieves the business objectives [15]. The tasks in this phase ensure that the data miner is heading for the right direction and getting closer to the business objectives, and thus vital to the whole data mining process.

In the context of the eMaintenance domain, prediction is a very common issue (such as failure prognosis) by resorting to data mining. Traditionally, a prediction model randomly samples from populations and measures performance in terms of mean error, either error rates or distance [74]. To evaluate a prediction model under the massive data circumstance, datasets are usually split into three subsets: training set, cross-validation set and testing set. How well the trained model fits to the testing set become the criterion to appraise the performance of a model. Since numerous studies have been done on evaluating performance of models [74], the details of which will not be discussed in this paper.

Generally, the obtained outcomes need to be reviewed and examined to check whether predetermined data mining objectives have been achieved. Moreover, as the range of maintenance data has been expanded to a broader scope, some unexpected yet interesting results might arise during the preceding phases. Therefore prior steps might need to be executed back-and-forth if necessary. Owing to the concrete tasks in this phase are highly dependent on the concerned data mining problem and could also vary from project to project, we will not go deep into other tasks in this phase.

3.6 Phase 6: Deployment

Creation of the model is generally not the end of a data-mining project. Even if the purpose of the model is to discover valuable knowledge from the chaos of data, the gained knowledge need to be organized and presented in a way that the end-user can easily perceive and make use of it [15].

To facilitate the deployment of data mining outcomes, data visualization tools are usually employed in this phase. Data visualization, also known as visual data exploration, is to present data in some visual form so that human can directly interact with data, draw conclusions and acquire insights into the data [75]. Data visualization plays an important role in the data mining process [76]. It enables direct involvement of domain experts and end-users and helps them understand the results intuitively and vividly. Besides, data visualization tools can also be used in the "data understanding" phase to conduct a preliminary exploration to data.

In the eMaintenance area, data visualization is also an indispensable element in supporting maintenance decision and even routine work. Such as, Augmented Reality (AR) applications were implemented in plants to assist maintenance engineers when performing maintenance tasks [77]. "Big maintenance data" have posed a challenge to traditional reporting tools or Business Intelligence (BI) systems that are no longer able to accommodate the scale, to capture the breathtaking changes and to fit to the variety types of Big Data. More superior data visualization solutions are needed to address the characteristics of Big Data and fulfil its' requirements.

A report [78] issued by The Data Warehouse Institute (TDWI) in 2011 indicates that Advanced Data Visualization (ADV) is a natural fit to Big Data analytics and has the greatest potential growth in the future. ADV is defined by Mohanty et al. [79] as a process whereby reliable data from multiple sources are amalgamated and visually communicated clearly and effectively through advanced graphical techniques. It enables analysts to present the data in a holistic, dynamic, multi-dimensional, diversified (such as hierarchies and neural nets) way, and hereby end-users could embrace a better cognitive understanding to the extracted knowledge. In another report [80], Forrester analysts Boris Evelson and Noel Yuhanna identified six key capabilities of ADV that could differentiate it with conventional data visualization solutions. The six capabilities to make data visualization tools comparatively advanced are dynamic data content, visual querying, multiple-dimension and linked visualization, animated visualization, personalization and business-actionable alerts.

As with the selection on data infrastructures and algorithms described in the previous sections, choosing an appropriate ADV tool relies on the demand of the practical application. The Forrester report [80] also proposed numerous criteria in terms of functional and technical capabilities of ADV to evaluate ADV platforms. The leading vendors and strong performers providing ADV solutions have been compared and listed in that report as well. Some of the industrial applications have already applied ADV to their eMaintenance solutions to support maintenance decision-making. For example, Blanes et al. [81] brought ADV into a fault detection and diagnosis system in the transportation industry.

Just as what Forrest [82] claimed: "It would be a mistake to slavishly follow the data miners to the point where you have lost the connection to reality, just like a driver should never blindly follow the GPS and ignore the reality outside the window", the greatest opportunity and risk lie in the last step in the path: applying the achieved patterns to business practice. Similarly, maintenance strategy can be optimized by modifying preventive maintenance intervals. Nevertheless, before applying these patterns to change operations, it is vital to examine the patterns which might do not make any sense and could lead to malfunction [37].

4. Challenges of Big Data mining in eMaintenance

Big data lead to not only a technology revolution, but also a business revolution, it provides an unprecedented opportunity for modern society and corporations [11]. The vast majority of value is hidden inside the maintenance data and it is like an iceberg floating in the sea waiting for further exploitation and utilization. Nevertheless, we are still facing huge challenges in mining the large-scale maintenance data.

Challenges of data mining in eMaintenance domain have been identified in previous research, such as, the complexity in selecting valuable datasets from multiple data sources [83]. From the Big Data point of view, the challenges consist of:

- integrating data from diverse sources

Industrial maintenance data with various formats are distributed over a multitude of sources. We can foresee that traditional

relational database management systems will play a role for quite a while in the commercial database market. How to effectively integrate data from RDBMS, non-relational databases, documents and other provenances is a big challenge.

- dealing with unstructured or semi-structured data

With limited or even no structure, a large proportion of the maintenance data are troublesome to process and analyse further, for example, it is difficult to conduct feature selection from the operational sound of a critical component when mingled with environmental noise.

- designing appropriate business models and algorithms

Over the last few decades, researchers have reaped the harvest of high-performance models and algorithms. However, multi-dimensional spare data, uncertain data and incomplete data can significantly distort the results of current algorithms. In addition, more advanced algorithms are required to tackle with unstructured or semi-structured data. Big Data have raised arduous requirements to algorithm design in terms of high scalability, error-tolerance, efficiency, and so forth.

Aside from above challenges, how to make data mining results easier to understand and act upon is yet another valuable and challenging topic. The research [84] indicates that the leading obstacle for organizations to extensively adopt analytics is lack of understanding to how to use analytics to improve the business rather than concerns with data and technology.

5. Conclusions

In this paper, we analysed the features of maintenance data, then proposed that maintenance data possess the three basic characteristics of Big Data: huge volume, rapid velocity and vast variety. Therefore traditional cognition to what maintenance data really are should be escalated to a higher and wider level. By reviewing existing literatures, we found that the attempts taken to utilize Big Data technologies so as to facilitate data mining in the eMaintenance field are limited both in number and scope. Subsequently, potential eMaintenance solutions, as summarized below, of applying Big Data technologies to maintenance data mining were explored in each phase of the CRISP-DM framework. Finally, challenges of Big Data mining in eMaintenance are discussed.

- Maintenance data storage

NoSQL data stores can play a significant role in storing humongous unstructured maintenance data with its highly scalable capacity and coping with the dynamically changed user requirements through its' schema-free pattern.

- Maintenance data analysis

Advancement in Artificial Intelligence (AI) and statistical approaches is paving the way for distilling useful information from maintenance data and supporting maintenance decision-making. Notably, some particular techniques like text mining, NLP, deep learning are now being intensively studied and may be employed to process unstructured or semi-structured data in the eMaintenance domain.

- Maintenance data computing

Distributed computing approaches like MapReduce model can be utilized to enhance the execution efficiency of complex algorithms when mining from huge amount of maintenance data.

- Maintenance data visualization

Advanced data visualization tools are able to provide a vivid and timely interaction between data and analysts, and hence efficiently transmit extracted patterns to end-users.

In summary, data mining in eMaintenance could be one of the most potential applications harnessing Big Data technologies. We are just at the initial stage of looking into these big maintenance data then extracting interesting and valuable patterns that lead to effective maintenance decision support. It is possible and imperative to leverage Big Data technologies to assist data mining for better eMaintenance support.

6. ACKNOWLEDGMENTS

We would like to thank all the reviewers for their insightful comments on this paper.

7. REFERENCES

- [1] X. Wu, X. Zhu, G. Wu, and W. Ding, "Data mining with big data," *Knowl. Data Eng. IEEE Trans.*, vol. 26, no. 1, 2014.
- [2] M. A. Beyer and D. Laney, "The Importance of 'Big Data': A Definition," Gartner, 2012.
- [3] Mira Kajko-Mattsson, Ramin Karim, Anita Mirjamdotter, "Essential Components of e-Maintenance," *Int. J. Performability Eng.*, vol. 7, no. 6, pp. 555–571, 2011.
- [4] R. Agrawal, A. Ailamaki, and P. Bernstein, "The Claremont report on database research," *Commun. ACM*, vol. 52, no. 6, pp. 56–65, 2008.
- [5] B. Babcock, S. Babu, and M. Datar, "Models and issues in data stream systems," in *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, 2002, pp. 1–16.
- [6] R. N. Romanowski, Carol J., "Methods, Analyzing maintenance data using data mining," in *Data mining for design and manufacturing*, Springer US, 2001, pp. 235–254.
- [7] M. Zaluski, S. Létourneau, J. Bird, and C. Yang, "Developing Data Mining-Based Prognostic Models for CF-18 Aircraft," *J. Eng. Gas Turbines Power*, vol. 133, no. 10, p. 101601, 2011.
- [8] R. Meseroll, C. Kirkos, and R. Shannon, "Data mining navy flight and maintenance data to affect repair," *Autotestcon*, 2007 IEEE, pp. 476–481, 2007.
- [9] D. Batanov, N. Nagarur, and P. Nitikhunkasem, "EXPERT-MM: A knowledge-based system for maintenance management," *Artif. Intell. Eng.*, no. 1993, pp. 283–291, 1993.
- [10] D. B. Malkoff, "A framework for real-time fault detection and diagnosis using temporal data," *Artif. Intell. Eng.*, vol. 2, no. 2, pp. 97–111, Apr. 1987.
- [11] V. Mayer-Schönberger and K. Cukier, *Big Data: A Revolution that Will Transform how We Live, Work, and Think*. Eamon Dolan/Houghton Mifflin Harcourt, 2013, p. 256.
- [12] A. Bahga and V. K. Madiseti, "Analyzing Massive Machine Maintenance Data in a Computing Cloud," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 10, pp. 1831–1843, Oct. 2012.
- [13] A. Rrays and T. Elematics, "Condition-based maintenance using sensor arrays and telematics," *Int. J. Mob. Netw. Commun. Telemat.*, vol. 3, no. 3, pp. 19–28, 2013.
- [14] A. Azevedo, "KDD, SEMMA and CRISP-DM: a parallel overview," *IADIS Eur. Conf. Data Min.*, pp. 182–185, 2008.
- [15] P. Chapman, J. Clinton, and R. Kerber, "CRISP-DM 1.0 Step-by-step data mining guide," 2000.
- [16] BSI, "Maintenance terminology," BS EN 133062001, 2001.
- [17] U. K. BSI, "Specification for the optimized management of physical assets," PAS 55-12008, 2008.
- [18] J. D. Fernandez and A. E. Fernandez, "Scada systems: vulnerabilities and remediation *," *J. Comput. Sci. Coll.*, vol. 20, no. 4, pp. 160–168, 2005.
- [19] I. E. Commission, "Functional safety – Safety instrumented systems for the process industry sector," IEC 61511-1, vol. Part 1, 2003.
- [20] T. Bangemann, X. Rebeuf, D. Reboul, A. Schulze, J. Szymanski, J.-P. Thomesse, M. Thron, and N. Zerhouni, "PROTEUS—Creating distributed maintenance systems through an integration platform," *Comput. Ind.*, vol. 57, no. 6, pp. 539–551, Aug. 2006.
- [21] R. Yu, B. Iung, and H. Panetto, "A multi-agents based E-maintenance system with case-based reasoning decision support," *Eng. Appl. Artif. Intell.*, vol. 16, no. 4, pp. 321–333, Jun. 2003.
- [22] A. Doan, J. F. Naughton, R. Ramakrishnan, A. Baid, X. Chai, F. Chen, T. Chen, E. Chu, P. Deroose, B. Gao, C. Gokhale, J. Huang, W. Shen, and B. Vuong, "Information Extraction Challenges in Managing Unstructured Data," *SIGMOD Rec.*, vol. 37, no. 4, pp. 14–20, 2008.
- [23] M. Y. Maarouf and S. M. Chung, "XML Integrated Environment for Service-Oriented Data Management," 2008 20th IEEE Int. Conf. Tools with Artif. Intell., pp. 361–368, Nov. 2008.
- [24] J. Pokorny, "NoSQL databases: a step to database scalability in web environment," *Int. J. Web Inf. Syst.*, vol. 9, no. 1, pp. 69–82, 2013.
- [25] C. Lynch, "How do your data grow?," *Nature*, vol. 455, 2008.
- [26] S. B. Fromm, Hansjörg, "Big data-Technologies and Potential," in *Enterprise-Integration*, Springer Berlin Heidelberg, 2014, pp. 107–124.
- [27] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Mag.*, pp. 37–54, 1996.
- [28] R. Wirth and J. Hipp, "CRISP-DM: Towards a standard process model for data mining," *Proc. Fourth Int. Conf. Pract. Appl. Knowl. Discov. Data Min.*, 2000.
- [29] T. R. Thomas Roth Berghofer, "MaMa: A Maintenance Manual for Case—Based Reasoning Systems," in *Case-Based Reasoning Research and Development*, 2001, pp. 452–466.

- [30] Bey-Temsamani, Abdellatif, Marc Engels, Andy Motten, Steve Vandenplas, Agusman P. Ompusunggu, "A practical approach to combine data mining and prognostics for improved predictive maintenance," *Data Min. Case Stud.*, vol. 36, 2009.
- [31] L. J. Koc M, "A system framework for next-generation e-maintenance system.," *China Mech. Eng.*, vol. 5, 2001.
- [32] R. Cattell, "Scalable SQL and NoSQL data stores," *ACM SIGMOD Rec.*, vol. 39, no. 4, 2011.
- [33] R. Hecht and S. Jablonski, "NoSQL evaluation: A use case oriented survey," 2011 Int. Conf. Cloud Serv. Comput. IEEE, pp. 336–341, Dec. 2011.
- [34] C. I. K. T.A.M.C. Thantriwate, "NoSQL query processing system for wireless ad-hoc and sensor networks," in *The International Conference on Advances in ICT for Emerging Regions*, IEEE, 2011, pp. 78–82.
- [35] H. Ito, M. Potekhin, and T. Wenaus, "Development of noSQL data storage for the ATLAS PanDA Monitoring System," *J. Phys. Conf. Ser.*, vol. 396, no. 5, p. 052041, Dec. 2012.
- [36] P. Vassiliadis, A. Simitsis, and S. Skiadopoulos, "Conceptual modeling for ETL processes," in *Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP.*, 2002, pp. 14–21.
- [37] P. Bastos, I. Lopes, and L. Pires, "A maintenance prediction system using data mining techniques," in *Proceedings of the World Congress on Engineering*, 2012, vol. III, pp. 2–7.
- [38] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark: cluster computing with working sets," in *Proceedings of the 2nd USENIX conference on Hot topics in cloud computing*, 2010, p. 10.
- [39] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The Hadoop Distributed File System," in *Mass Storage Systems and Technologies (MSST)*, 2010 IEEE 26th Symposium on, 2010, pp. 1–10.
- [40] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107–113, 2008.
- [41] M. Lee, H. Lu, T. Ling, and Y. Ko, "Cleansing data for mining and warehousing," *Database Expert Syst. Appl.*, pp. 751–760, 1999.
- [42] R. Pearson, "Exploring process data," *J. Process Control*, vol. 11, pp. 179–194, 2001.
- [43] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, "Methodologies for data quality assessment and improvement," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–52, Jul. 2009.
- [44] M. Al-Jumaili and V. Rauhalu, "Data quality in eMaintenance: a call for research," in *International Conference on Maintenance Performance Measurement & Management: MPM2*, 2011, p. 69.
- [45] T. H. Kyle Banker, Peter Bakkum, "MongoDB in action," 2nd ed., Manning publications Co., 2013, p. 375.
- [46] Dorian Pyle, "Data preparation for data mining," morgan kaufmann, 1999.
- [47] J. Cortiñas Abrahantes, C. Sotto, G. Molenberghs, G. Vromman, and B. Bierincx, "A comparison of various software tools for dealing with missing data via imputation," *J. Stat. Comput. Simul.*, vol. 81, no. 11, pp. 1653–1675, Nov. 2011.
- [48] J. M. Jerez, I. Molina, P. J. García-Laencina, E. Alba, N. Ribelles, M. Martín, and L. Franco, "Missing data imputation using statistical and machine learning methods in a real breast cancer problem," *Artif. Intell. Med.*, vol. 50, no. 2, pp. 105–115, Oct. 2010.
- [49] J. Hellerstein, "Quantitative data cleaning for large databases," *United Nations Econ. Comm. Eur.*, pp. 1–42, 2008.
- [50] K. Beyer and J. Goldstein, "When is 'nearest neighbor' meaningful?," *Database Theory—ICDT'99*, pp. 217–235, 1999.
- [51] H. Kriegel, P. Kröger, E. Schubert, and A. Zimek, "Outlier detection in axis-parallel subspaces of high dimensional data," *Adv. Knowl. Discov. Data Mining.*, vol. 1, pp. 831–838, 2009.
- [52] H.-P. Kriegel, M. S. Hubert, and A. Zimek, "Angle-based outlier detection in high-dimensional data," *Proceeding 14th ACM SIGKDD Int. Conf. Knowl. Discov. data Min.*, pp. 444–452, 2008.
- [53] Y. Bassil, "A Survey on Information Retrieval, Text Categorization, and Web Crawling," *J. Comput. Sci. Res.*, vol. 1, no. 6, pp. 1–11, 2012.
- [54] W. Zou, A. Ng, S. Zhu, and K. Yu, "Deep Learning of Invariant Features via Simulated Fixations in Video.," *NIPS*, pp. 3212–3220, 2012.
- [55] M. Jiu, C. Wolf, G. Taylor, and A. Baskurt, "Human body part estimation from depth images via spatially-constrained deep learning," *Pattern Recognit. Lett.*, pp. 1–8, Oct. 2013.
- [56] J. Han, M. Kamber, and J. Pei, "Data mining: concepts and techniques," 2nd ed., Diane Cerra, 2006, pp. 70–72.
- [57] S. Kandel, A. Paepcke, J. Hellerstein, and J. Heer, "Wrangler: Interactive visual specification of data transformation scripts," in *PART 5-Proceedings of the 2011 annual conference on Human factors in computing systems*, 2011, pp. 3363–3372.
- [58] Z. Prekopcsák, G. Makrai, H. Tamas, and C. Gaspar-Papanek, "Radoop: Analyzing big data with rapidminer and hadoop," in *Proceedings of the 2nd RapidMiner Community Meeting and Conference (RCOMM 2011)*, 2011, pp. 1–12.
- [59] S. Santini and K. Romer, "An adaptive strategy for quality-based data reduction in wireless sensor networks," in *Proceedings of the 3rd International Conference on Networked Sensing Systems (INSS 2006)*, 2006, pp. 29–36.
- [60] C. Park, "Aggressive Data Reduction for Damage Detection in Structural Health Monitoring," *Struct. Heal. Monit.*, vol. 9, no. 1, pp. 59–74, Jun. 2010.
- [61] G. G. EJ Msechu, "Sensor-Centric Data Reduction for Estimation With WSNs via Censoring and Quantization," *Signal Process. IEEE Trans.*, vol. 60, no. 1, pp. 400–414, 2012.

- [62] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. John Wiley & Sons, 2012.
- [63] T. Duda and M. Canty, "Unsupervised classification of satellite imagery: Choosing a good algorithm," *Int. J. Remote Sens.*, vol. 23, no. 11, pp. 2193–2212, Jan. 2002.
- [64] E. Elbeltagi, T. Hegazy, and D. Grierson, "Comparison among five evolutionary-based optimization algorithms," *Adv. Eng. Informatics*, vol. 19, no. 1, pp. 43–53, Jan. 2005.
- [65] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 161–168.
- [66] M. Mazid, S. Ali, and K. Tickle, "Improved C4. 5 algorithm for rule based classification," *Proc. 9th WSEAS Int. Conf. Artif. Intell. Knowl. Eng. data bases*, pp. 296–301, 2010.
- [67] L. Hall, K. Bowyer, and W. Kegelmeyer, "Distributed learning on very large data sets," in *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2000, pp. 79–84.
- [68] S. Yoo, M. Harman, and S. Ur, "Highly scalable multi objective test suite minimisation using graphics cards," *Search Based Softw. Eng.*, pp. 219–236, 2011.
- [69] C. Chu, S. Kim, Y. Lin, Y. Yu, and G. Bradski, "Map-reduce for machine learning on multicore," *NIPS*, vol. 6, pp. 281–288, 2006.
- [70] G. Wu, H. Li, X. Hu, and Y. Bi, "MRc4. 5: C4. 5 ensemble classification with MapReduce," *ChinaGrid Annu. Conf. 2009*, IEEE, pp. 249–255, 2009.
- [71] L. Di Geronimo, F. Ferrucci, A. Murolo, and F. Sarro, "A Parallel Genetic Algorithm Based on Hadoop MapReduce for the Automatic Generation of JUnit Test Suites," *Softw. Testing, Verif. Valid. (ICST)*, 2012 IEEE Fifth Int. Conf., pp. 785–793, Apr. 2012.
- [72] Y. He, H. Tan, W. Luo, S. Feng, and J. Fan, "MR-DBSCAN : a scalable MapReduce-based DBSCAN algorithm," *Front. Comput. Sci.*, vol. 8, no. 1, pp. 83–99, 2014.
- [73] R. Polikar, "Ensemble learning," in *Ensemble Machine Learning*, Springer US, 2012, pp. 1–34.
- [74] N. I. Sholom M. Weiss, "Predictive Data Mining: A Practical Guide," 1st ed., morgan kaufmann, 1998, p. 228.
- [75] D. Keim, "Information visualization and visual data mining," *Vis. Comput. Graph. IEEE Trans.*, vol. 8, no. 1, pp. 1–8, 2002.
- [76] M. De Oliveira and H. Levkowitz, "From visual data exploration to visual data mining: A survey," *Vis. Comput. Graph. IEEE Trans.*, vol. 9, no. 3, pp. 378–394, 2003.
- [77] S. K. Ong, M. L. Yuan, and a. Y. C. Nee, "Augmented reality applications in manufacturing: a survey," *Int. J. Prod. Res.*, vol. 46, no. 10, pp. 2707–2742, May 2008.
- [78] P. Russom, "Big data analytics," 2011.
- [79] H. S. S Mohanty, M Jagadeesh, "Big Data Imperatives: Enterprise 'Big Data' Warehouse, 'BI' Implementations and Analytics," Aspress, 2013.
- [80] B. Evelson and K. TaKeaWays, "The Forrester Wave™: Advanced Data Visualization (ADV) Platforms, Q3 2012," 2012.
- [81] L. Blanes, A. Costa, N. Réhault, and M. Keane, "Integration of Fault Detection and Diagnosis with Energy Management Standard ISO 50001 and Operations and Maintenance of HVAC Systems," 2013.
- [82] F. Shull, "Getting an Intuition for Big Data," *Software, IEEE*, vol. 30, no. 4, pp. 3–6, Jul. 2013.
- [83] R. Karim, P. Söderholm, and O. Candell, "Development of ICT-based maintenance support services," *J. Qual. Maint. Eng.*, vol. 15, no. 2, pp. 127–150, 2009.
- [84] S. Lavalle, E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz, "Big Data , Analytics and the Path From Insights to Value," *MIT Slogan Management Review*, no. 52205, 2011.



Luleå Railway Research Center

JVTC is the best platform for research and innovation in reliability and maintenance of the railway system.

Dependability - RAMS, LCC, Risk Analysis

Information Logistics - eMaintenance Solutions

Wear and Failure Reduction

Maintenance Management and Human Factors

Whenever you think of research and innovation in railway maintenance, think of us!

www.jvtc.ltu.se



CMIS

Center for Maintenance and Industrial services

CMIS is a competence center within Luleå University of Technology in the field of maintenance, which encompasses research, training and dissemination of information.

eMaintenance & Information Logistics

RAMS

Remaining Useful Life Estimation

LCC/LCP

Maintenance Decision Modelling

Condition Monitoring

Diagnostic & Prognostic

Human Factors

www.cmis.ltu.se

L
LULEÅ
UNIVERSITY
OF TECHNOLOGY

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. Phillip Tretten, Organizing chair

www.emaintenance2014.org

ISBN 978-91-7439-972-1 (print)

ISBN 978-91-7439-973-8 (pdf)