



Proceedings of the 9th IMA International Conference on Modelling
in Industrial Maintenance and Reliability
London, UK
12-14 July, 2016

Edited by:
Philip Scarf, Shaomin Wu, Phuc Do

ISBN: 978-0-905091-31-0

Table of contents

- 1 Risk based maintenance of railway rolling stock assets incorporating competing risks models
B. Alkali and F. Dinmohammadi
- 8 A contribution to spare parts management based on a dynamic model for stock replenishment
P. Almeida, E. Henriques and V. Infante
- 14 Optimising life extension and maintenance policy decisions for safety critical elements
I. Animah, M. Shafiee and N. Simms
- 18 Modelling on joint optimization for two-dimensional age replacement interval and spare parts provisioning
Y. Bai, H. Qiwei and Z. Jianmin
- 29 Calendar-based age replacement policy and its generalizations
D. Banjevic, M. A. Bajestani
- 35 Parametric bootstrap goodness-of-fit tests for imperfect maintenance models
C. Chauvel, J.Y. Dauxois, L. Doyen and O. Gaudoin
- 41 Residual life estimation based on multi-phase nonlinear wiener process
H. Chen, B. Guo
- 47 Building a multicriteria decision model for risk evaluation in gas pipelines
A. T. de Almeida, M. H. Alencar and C. Medeiros
- 52 Reliability assessment of buried pipelines under seismic action
A.U. Eбенуwa and K.F. Tee
- 58 Carriageway deterioration prognosis modelling using Markovian chain
E. O. Ekpiwhre, K.F. Tee, O. Mordi and T. Bull
- 64 A data quality metrics hierarchy for reliability data
R. Gitzel, S. Turrin, S. Maczey, S. Wu and B. Schmitz
- 70 A two-phase life prediction model of explosion proof valve of lithium ion battery under continuous variable-stress
W. Gong, Y. Chen, W. Zhang and R. Kang
- 77 Maintenance optimization for systems with non-stationary degradation and random shocks
C. Guo, Y. Bai and Y. Jia
- 84 An extended log-linear model for modeling wear processes
R. Jiang
- 90 Improved preventive maintenance of a rotating machine by fault monitoring of its components via a neural network model
S. Khamel and Kh. Bouacha
- 96 Modelling asset management of railway overhead line equipment
P. Kilsby, R. Remenyte-Prescott and J.D. Andrews
- 102 Condition-based maintenance at both scheduled and unscheduled opportunities
Sz. Kyalosi, S. Kapodistria and J. A. C. Resing
- 108 Competing deterioration processes modeling and RUL estimation based on factorial hidden Markov models
T. T. Le, F. Chatelain and C. Bérenguer
- 114 Life model of aluminum Alloy crevice corrosion based on mechanical property under coupling environment conditions
Z. Li, P. Lei, F. Bai and J. Wang
- 121 Condition-based maintenance for degrading systems with state-dependent operating cost
B. Liu, M. Xie and W. Kuo
- 127 Using observers and bond graph in modelling and multiple fault identification
B. Maaref, H. Dhouibi, Z. Simeu-Abazi and H. Messaoud
- 133 Agent-based Maintenance System Methodology for optimal Reconfigurable Vibrating Screen Management
O. A. Makinde, K. Mpofu, B. I. Ramatsetse and M. K. Adeyeri
- 142 Robotic system methodology for automated maintenance of a reconfigurable vibrating screen
O. A. Makinde, K. Mpofu, B. I. Ramatsetse and M. K. Adeyeri
- 150 Modelling imperfect repair in a degradation system: a stochastic comparison
S. Mercier and I. T. Castro

- 156 Shock and damage model parameter estimation and forecasting with an unobserved number of shocks
N. Montgomery and D. Banjevic
- 162 A model reduction for probabilistic-reliability (RPR) static analysis using a set of representative functions
L. Ncib, A. Manai, H. Hassis and N. Gmati
- 168 Adjusting condition monitoring quality for condition-based maintenance optimization of deteriorating systems
T. P. K. Nguyen, K. T. Huynh, P. Do, C. Bérenguer and A. Grall
- 174 A method of dependent failure-occurrence detection for parallel systems
S. Ota and M. Kimura
- 179 RCM cost analysis and simulation of MV XLPE cable failure process towards optimum preventive maintenance strategy
M. Qatan, B.M. Alkali, M. Farrag, and C. Zhou
- 186 Modelling of operational reliability and maintenance cost for avionics systems with permanent and intermittent failures
A. Raza and V. Ulansky
- 193 Research on the application of association rules in fault diagnosis of high-speed railway in China
J. H. Ren, F. Liu and H. Hu
- 199 Joint optimization of maintenance preventive "Piggyback" and spare parts policies
T. F. O. Rodriguez, L. F. A. Ribeiro and C. A. V. Cavalcante
- 205 Buffer allocation for continuous imperfect production lines by simulation and analytic network process
O. M. Topcu and S. I. Satoglu
- 211 Redundancy allocation for series-parallel warm-standby systems with interval-valued lifetime deceleration factor
W. Wang, J. Xiong and M. Xie
- 217 Predicting discounted warranty cost over Time considering failed-but-not-reported events
X. Wang, Z. Ye, W. Xie and L.C. Tang
- 224 On assumptions in optimisation of warranty policies
S. Wu, R. Gitzel and S. Turrin
- 230 Mission reliability evaluation of repairable PMS with Bayesian networks
X. Wu
- 235 Inspection and maintenance optimization of a multi-state series-parallel system
M. A. Yin, Z. L. Sun, J. Wang and Y. Guo
- 242 A new modelling methodology for the assessment of maintenance policies of subsea systems
Y. Zhang, A. Rauzy and A. Barros
- 248 Analysis of risk events in UAV flight test project based on Bayesian network
S. Zhenfei, L. Baocheng and M. Lin

255 **Index of Authors**

Risk based maintenance of railway rolling stock assets incorporating competing risks models

B. Alkali* and F. Dinmohammadi**

*, ** *Department of Engineering, School of Engineering and Built Environment, Glasgow Caledonian University, G4 0BA: babakalli.alkali@gcu.ac.uk, fateme.dinmohammadi@gcu.ac.uk*

Abstract: This paper addresses some of the problems associated with failure analysis and maintenance management of a fleet of Electric Multiple Unit (EMU) trains in Scotland. The Class 380 EMUs are identified as significant service fleet operating on busy routes in Scotland. Safety critical systems that could possibly contribute to the fleet in-service failure is considered in this paper. The failure modes of the critical systems in terms of risk of failure are identified. The competing risk model is proposed in an attempt to identify the marginal distribution from the competing risk data. It is well known that the marginal distributions of the competing risk variables are identifiable if we assume independence. Failure of any of the critical systems on the train occurs randomly due to a specific category of failure modes which is assumed to follow an exponential distribution. We discuss the competing risk model and generate useable data at maintenance censoring and censoring from multiple failure modes and further go on to propose the random exponential model as a maintenance optimization model.

Keywords: Risk-based maintenance, railway transport, rolling stock, competing risk, failure analysis.

1. Introduction

Rolling Stock in Scotland is primarily operated by Abellio Scotrail and operates 95% of passenger rail services in Scotland, Abellio Scotrail provide approximately 2000 services operations a day. The Class 380 EMU fleet runs primarily on Ayrshire, Inverclyde and Renfrewshire routes, boosting Scotland's transport system by adding more than 7,500 passenger seats to the rail network. There are 38 Class 380 EMU which consist of three and four car configuration the 380 EMU were introduced into service in September 2010 delivering substantial benefits for passengers across the ScotRail network in terms of capacity, safety and customer comfort. Each unit has spacious seating, wide aisles, air-conditioning, power sockets for laptops and handheld devices under each table, luggage provision, space for cycles and wheelchairs, plus CCTV. The Class 380 monitors its own energy use and has regenerative braking that can return energy to the overhead line power system for use by other trains.

In this paper we attempt to use the competing risk framework to effectively and efficiently optimise maintenance on the Class 380 rolling stock by analysing the current maintenance strategy against system failure trends to date and also adequate maintenance philosophies in order to improve reliability and integrity of the Class 380 train units. The performance of rolling stock is always threatened by reliability. The failure of critical systems such as wheel set, doors, pantograph and the couplers amongst others can lead to service delays, cancellation of service, and hence reduced overall performance of the entire fleet. The risk of failure in relation to reliability performance is considered within the competing risk framework.

2. The competing risk framework

Competing risk models essentially describe a situation in which the minimum of two or more lifetime variables is observed together with the indicator of which variable is the smallest. The unobserved variable is said to be censored by the observed variable. The theory of competing risk requires the assumption that the system is repaired to "as good as new" (or sufficiently closely to this) so that we can consider the reliability data as generated by a sequence of independent and identically distributed observations. For each observation we have a competing risk model.

Let $Y_i=1, 2, \dots$ be independent copies of Y . The process $Y=Y_1, Y_2, \dots$ is called the competing renewal process associated with Y . From the observable data we are able to estimate two basic quantities called the sub-survivor functions in equation (1):

$$S_x^*(t) = P(X > t, X < Z) \text{ and } S_z^*(t) = P(Z > t, X > Z) \quad (1)$$

However, we are unable to estimate the survivor functions for X and Z in equation (2), unless we make non-testable assumptions

$$S_z(t) = P(Z > t) \quad \text{and} \quad S_x(t) = P(X > t) \quad (2)$$

The most common such assumption is that the variables are independent. In this case the distributions of X and Z are identifiable Tsiatis (1975), Crowder (1991) and we are able to estimate the survivor functions using the Kaplan-Meier estimator. (The main practical problem in reliability data is that identifiability only holds up to the essential supremum of Y , so where for example equipment is replaced after a fixed period we are not able to estimate the lifetime after that period without making distributional assumptions that allow us to interpolate outside the dataset). However, in many reliability situations the assumption of independence may be considered, and as a whole range of different joint distributions for (X, Z) may be compatible with a given pair of sub-survivor functions and hence we are forced to make non-testable assumptions when applying competing risk models to interpret operating data. For further discussion of competing risk models see (Tsiatis 1975; Crowder 1991; Crowder 1994; Bunea and Mazzuchi 2006; Cooke and Morales-Mapoles 2006; Langseth and Lindqvist 2006).

3. Competing risk models

Competing risk models essentially describe a situation in which the minimum of two or more lifetime variables is observed; together with the indicator of which variable is the smallest. The unobserved variable is said to be censored by the observed variable. Applications of the classical competing risk model assume renewals of the equipment, that is, the equipment is restored to as good as new. The application of the competing risk concept in the railway industry is scarce and the main contribution of this paper is to address the opportunity of applying the proposed model to the class 380 EMU train units.

It is not necessary to assume that the different variables in the competing risk are independent. Indeed, it has been argued that for reliability applications this is not necessarily the case. In particular, failure mode failure times may be statistically dependent because of coupled degradation processes. Failure times and preventive maintenance times may be dependent because PM is conducted taking some account of the potential failure time, or partial knowledge of the state of the equipment.

The competing risk problem is that of identifying the underlying (marginal) distributions from the operational data. In general we are unable to identify the marginals from the data without making untestable assumptions (Bedford and Meilijson 1997; Aven and Jensen 1999; Crowder 2001). The motivation for studying this problem is that we would like to know how things would change if we were able to abolish a failure cause, stop doing maintenance, or perform some other action that changes the underlying relationship of the variables. From an operational research point of view, the competing risk problem is one example of a the more general modelling problem in which we need to construct models that support decision making, but cannot always easily identify those models from existing data.

Various competing risk models in reliability have been developed (see for example the overview given in (Bedford 2005)). Some of these models explicitly try to address the relationship between failure and PM. For example the random signs model of (Cooke 1993) suggests that PM time is randomly either larger or smaller than failure time independently of the actual failure time, thus modelling a situation in which a maintainer aims to perform PM just before the actual failure time. A more general version of this model is that described in (Lindqvist, Stove *et al.* 2006) which uses a "repair alert function" that describes the likelihood of the crew to carry out the repair at any point in time prior to the failure time. Failure data from any real system is of course highly censored, and one of the motivations in writing this paper is to build up a better picture of the

way in which failure data is censored. While competing risk models do not have to involve preventive maintenance as a censoring mechanism, it is clear that this involves an important category of such models.

The construction of competing risk models involving maintenance is closely allied to the construction of maintenance optimization models. However, in some sense the two classes of models approach the problem from different ends. For competing risk models, we are trying to make sense of a censoring structure that is already present in the data: simple assumptions are sought which would enable us to identify underlying failure rates given the data. For maintenance optimization models we search through a decision space parameterized by some natural decision variable, to optimize some heuristic measure of maintenance efficiency. Considerable thought has been put into the development of such models and they present us with broad sets of approaches to model existing maintenance policies even when these are not wholly thought out or optimized. It is worth therefore spending a little time discussing the relationship with maintenance optimization models.

4. Maintenance optimization models

Many different types of maintenance optimization models have been developed and examples of such models can be found in (Dekker 1996; Wang 1997; Dekker and Scarf 1998; Wang 2002). (Wang 2002) for example gives a nice overview of maintenance models distinguishing such policies as age replacement, block replacement, failure limit, repair limit, sequential, and repair counting; optimization criteria such as cost, availability, failure rate limits, cost and reliability; different time horizons over which the criteria are measured, and different types of repair. It is acknowledged that one of the reasons that many models have found limited application is that the assumptions made are typically rather strong. For example, many models assume just a single component in isolation and this has led to the proposed random exponential model discussed in Section 5.

5. Random exponential model

The random exponential model proposed was thought through from careful consultation with rolling stock maintenance and continuous improvement engineers. It became clear to the authors that not all critical systems and extraneous information associated to the train in service are recorded. There are lots of unrecorded features which are largely impossible to determine retrospectively. However, trains can go out of service for many different reasons, if for example we take the door systems. Apart from the internal failure of the door components, the doors can be taken out of service due to failure of other equipment within the functional boundary or doors during service. To capture this kind of behavior the following random exponential type has been proposed. The time to failure of the door internal component is the smallest of a number of independent exponentially distributed failure times. However, because some factors are only sometimes present – for example the process failures – we model this by factors that are randomly present. To some up, the up time is modelled as

$$T = \min(T_1, T_2, \dots, T_n)$$

where T_i is exponentially distributed with failure rate λ_i with probability p_i , and equals ∞ with probability $q_i = 1 - p_i$, $i = 1, \dots, n$. Note that, unless some $p_i = 1$, there is a positive probability that all of the T_i are infinite, implying that T has a defective distribution. In practical terms this corresponds to a situation in which equipment can run for a long time when “problem generating conditions” are not present. Of course the mean of T is infinite in this model, but in practice one is only ever interested in considering a bounded time domain anyway, so this is not a relevant parameter for the practical situation. The T_i are assumed mutually independent. Hence this is an independent competing risk model, but with degenerate risks (that is, the distributions of the risks are not all proper). A little thought shows that the distribution of T (conditional on it being finite) is a mixture of exponentials:

T has failure rate $\sum_{i=j_1, \dots, j_r} \lambda_i$ with probability $\prod_{i=j_1, \dots, j_r} p_i \prod_{i \neq j_1, \dots, j_r} q_i$, for each set of integers $j_1, \dots, j_r$.

In the setting we describe it is natural to have the competing risk data available (and in our particular application this is the case in principle, though it has to be assembled at some cost from a variety of different data sources). The competing risk data is data of the form T, I where T is the time at which the equipment goes down, and I is the cause of the equipment going down. This data enables us to estimate the sub-survivor functions (defined below) and also the probability that $T = \infty$ (in practice, that the failure was not observed in our time window). Using the expression *h.o.t.* for higher order terms, that is, exponentials converging to 0 more rapidly than the given term as $t \rightarrow \infty$, we have

$$\begin{aligned} P(T_i = t, T_j > t, j \neq i) &= p_i \lambda_i \exp(-\lambda_i t) \prod_{j \neq i} (q_j + p_j \exp(-\lambda_j t)) \\ &= p_i \lambda_i q_1 \dots q_{i-1} q_{i+1} \dots q_n \exp(-\lambda_i t) + h.o.t. \end{aligned} \quad (1)$$

From this we see that the sub-survivor functions, $S_i^*(t) = P(T_i > t, T_j > T_i, j \neq i)$ satisfy

$$\begin{aligned} S_i^*(t) &= P(T_i > t, T_j > T_i, j \neq i) = \int_t^\infty p_i \lambda_i \exp(-\lambda_i s) \prod_{j \neq i} (q_j + p_j \exp(-\lambda_j s)) ds \\ &= \int_t^\infty p_i \lambda_i \exp(-\lambda_i s) \sum_{\substack{i_j=0,1 \\ i \neq j}} q_1^{1-i_1} \dots q_n^{1-i_n} p_1^{i_1} \dots p_n^{i_n} \exp(-\sum_j i_j \lambda_j s) ds \\ &= \sum_{\substack{i_j=0,1 \\ i \neq j}} \frac{p_i \lambda_i q_1^{1-i_1} \dots q_n^{1-i_n} p_1^{i_1} \dots p_n^{i_n}}{\lambda_i + \sum_j i_j \lambda_j} \exp(-(\lambda_i + \sum_j i_j \lambda_j)t). \end{aligned} \quad (2)$$

The sub-survivor functions can be estimated from competing risk data. In competing risk problems we are interested to know whether the underlying model is identifiable. This means, if we know the sub-survivor functions (which are in principle observable), are the model parameters uniquely determined? The contra positive question is equivalent to asking: Suppose two sets of parameters are given $(p_1, \dots, p_n, \lambda_1, \dots, \lambda_n)$ and $(p'_1, \dots, p'_n, \lambda'_1, \dots, \lambda'_n)$. Can these give the same sub-survivor functions? We shall show below that the model class is identifiable, and give a way of determining the parameters.

Identification of the rates

The values of the rates λ_i can clearly be identified. The expression above shows that the sub-survivor function $S_i^*(t)$ is dominated by the $\exp(-\lambda_i t)$ term for large t . Hence we can identify λ_i .

Identification of the probabilities

We note first that by considering the dominant terms in the sub-survivor functions at some large t we can obtain a number of equations just involving the p_i . For each $S_i^*(t)$ the dominant term in the summation is the one where $i_1, \dots, i_n = 0$ and the corresponding coefficient (ignoring the terms only involving the λ_j is $p_i \prod_{j \neq i} q_j$. Hence, for each i , we can determine (to any required degree of accuracy) the value of $p_i \prod_{j \neq i} q_j$. In order to identify the parameters we would need to solve these n non-linear equations for the p_i . Solving directly is not a good strategy though, as the numerical solutions converge poorly. Much better is to transform into a fixed point problem. As stated above in equation (4), using the sub-survivor functions we can obtain a set of equations

$$\alpha_i = p_i \prod_{j \neq i} q_j = \frac{1 - q_i}{q_i} \prod_{j=1}^n q_j \quad \text{for all } i,$$

where the α_i are known. If we knew $x = \prod_{j=1}^n q_j$ then we could solve for $q_i = \frac{x}{\alpha_i + x}$. However, x is the probability that T is infinite, which is also part of the competing risk data (in practice, this is the probability that T was not observed in our time window). Hence the probabilities are also identifiable. This model enables us to capture the effect of different sorts of risks of failure of the door critical system, and in

particular, to be able to model the effect of removing different sorts of potential failures. Hence it can support decision making for the risk of failure reduction.

6. Discussions

The main assumption of the proposed model is that T_i are drawn randomly from an exponential distribution and this is somewhat similar to the mixture of exponential model by Bunea, Mazzuchi et al. (2006). From the data we can observe what causes the doors to fail and what prevent doors from functioning. The actual data consists of a sequence of failure modes which are followed by either maintenance or a failure. The model presented here has independent censoring, in contrast to many of the other competing risk models developed in the reliability context. In principle, since the underlying distributions are always identifiable when we have independent censoring, the identifiability issue is not the problem as with other models. One important piece of information would be available in addition to the competing risk data is the rate at which failure arise and this could be identified from the data collected. The sub-functions can in principle be estimated from the observable data. Clearly, the observable data is given by a competing risk model.

The model is slightly restrictive in allowing for a degenerate lifetime distribution. However, this reflects a practical issue that is nowadays quite often the case: Underlying systems reliability may be measured during maintenance. However, during operation there are lots of things, often external to the door system itself, which can cause the train to be taken out of service. In this way it is possible to have a high measured reliability during maintenance times and a much lower availability in practice. Judgmental information is also required to model how changes to current set up would affect availability of the trains. Investigation into the dataset collected is still underway and we intend to fit the proposed model and present results in future publications.

Acknowledgements

We gratefully acknowledge the support and assistance of Abellio Scotrail. We thank Abhinay Ramani for his cooperation in this work.

References

- Alkali, B., T. Bedford and J. Gaw. 2006. Failure and Maintenance Data Extraction from Power Plant Maintenance Management Databases. ALT Conference on Degradation, Damage, Fatigue and Accelerated Life Models in Reliability Testing. Angers, France.: 329-334.
- Alkali, B., T. Bedford, J. Quigley and J. Gaw. 2007. Failure and Maintenance Data Extraction from Power Plant Maintenance Databases. to appear in the special issue in Journal of Statistical Planning and Inference.
- Aven, T. and R. Dekker. 1997. A useful framework for optimal replacement models. Reliability Engineering & System Safety **58**(1): 61-67.
- Bedford, T., 2004. Assessing the Impact of Preventive Maintenance Based on Censored Data. Quality and Reliability Engineering International **20**(3): 247-254.
- Bedford, T. and B. H. Lindqvist. 2004. The identifiability problem for repairable systems subject to competing risks. Advances in Applied Probability **36**(3): 774-90.
- Bunea, C. and T. Mazzuchi. 2006. Competing Failure Modes in Accelerated life Testing. Journal of Statistical Planning and Inference **136**(5): 1608-1620.
- Bunea, C., T. Mazzuchi, S. Sarkani and H. Chang. 2006. Application of modern reliability database techniques to military system data. Reliability Engineering and System Safety(doi:10.1016/j.res.2006.10.011).

- Christer, A. H. and W. M. Waller. 1984. Operational Research Approach to Planned Maintenance: Modelling PM for a Vehicle Fleet. *Journal of the Operational Research Society* **35**: 967-984.
- Cooke, R. and O. Morales-Mapoles. 2006. Competing Risk and the Cox Proportional Hazards Model. *Journal of Statistical Planning and Inference* **136**(5): 1621-1637.
- Cooke, R. M., 1996. Design of reliability data bases, Part II: competing risk and data compression. *Reliability Engineering and System Safety* **51**(2): 209-223.
- Crocker, J. and U. D. Kumar. 2000. Age-related maintenance versus reliability centred maintenance: a case study on aero-engines. *Reliability Engineering & System Safety* **67**(2): 113-118.
- Crowder, M., 1991. On the identifiability crisis in competing risks analysis. *Scandinavian Journal of Statistics* **18**: 223-233.
- Crowder, M., 1994. Identifiability crises in competing risks. *International Statistical Review* **62**: 379-391.
- Dagpunar, J. S., 1996. A maintenance model with opportunities and interrupt replacement options. *Journal of the Operational Research Society* **47**(11): 1406-1409.
- Dekker, R., 1996. Applications of maintenance optimization models: A review and analysis. *Reliability Engineering & System Safety* **51**(3): 229-240.
- Dekker, R. and M. C. Dijkstra. 1992. Opportunity-Based Age Replacement - Exponentially Distributed Times between Opportunities. *Naval Research Logistics* **39**(2): 175-190.
- Dekker, R. and P. A. Scarf. 1998. On the impact of optimisation models in maintenance decision making: the state of the art. *Reliability Engineering & System Safety* **60**(2): 111-119.
- Dekker, R. and E. Smeitink. 1991. Opportunity-Based Block Replacement. *European Journal of Operational Research* **53**(1): 46-63.
- Guess, F. and F. Proschan. 1988. Mean residual life, theory and applications. *Handbook of Statistics Reliability and Quality Control* **7**: 214-224.
- Jhang, J. P. and S. H. Sheu. 1999. Opportunity-based age replacement policy with minimal repair. *Reliability Engineering & System Safety* **64**(3): 339-344.
- Langseth, H. and B. Lindqvist. 2006. Competing Risks for Repairable Systems: A Data Study. *Journal of Statistical Planning and Inference* **136**(5): 1687-1700.
- Lindqvist, B. and K. Doksum. 2003. *Mathematical and Statistical Methods in Reliability. Series on Quality, Reliability and Engineering Statistics. UK, Vol 7, World Scientific.*
- Lindqvist, B., B. Støve and H. Langseth. 2006. Modelling of dependence between critical failure and preventive maintenance: The repair alert model. *Journal of Statistical Planning and Inference* **136**(5): 1701-1717.
- Nanda, A., S. Bhattacharjee and S. Alam. 2006. Properties of proportional mean residual life model. *Statistical and Probability letters* **76**(9): 880-890.
- Satow, T. and S. Osaki. 2003. Opportunity-based age replacement with different intensity rates. *Mathematical and Computer Modelling* **38**(11-13): 1419-1426.

Tsiatis, A., 1975. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences of the USA* **72**(1): 20-22.

Wels, H., J. Aydt, R. Kirsch and H. Meier. 2006. VGB Database for the forced unavailability of power plants. *Safety and Reliability for Managing Risk - Guedes Soares and Zio (eds)*: 985-989.

Xie, M., T. Goh and Y. Tang. 2004. On changing points of mean residual life and failure rate function for some generalized Weibull distributions. *Reliability Engineering and System Safety* **84**(3): 293-299.

A contribution to spare parts management based on a dynamic model for stock replenishment

P. Almeida¹, E. Henriques^{2*}, V. Infante²

¹ Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049-001 Lisboa, Portugal

² LAETA, IDMEC, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049-001 Lisboa, Portugal

*Corresponding author. E-mail address: elsa.h@tecnico.ulisboa.pt

Abstract. Spare parts management plays a major role in the efficiency and perceived value of maintenance services for automotive vehicles. Spare parts have to be available to guarantee an adequate level of service. However, a negative impact on the indirect cost of the services is expected if an excessive quantity of spares is kept available, meaning that a trade-off has to be made. This paper presents an operational model to improve the management of spares for maintenance providers, considering aims of low costs of stock and high service levels. To deal with the different demand patterns, the spares are classified based on demand variability and on the mean interval between consecutive demands, allowing the selection of appropriated forecast methods to estimate the next period demand. It is then proposed to manage the replenishment of stock based on a dynamic reorder point and order quantity calculated considering at each time period the updated demand patterns and error measures of past estimates. The model was tested in a real case study. The results show an improvement on the global service level, essentially due to a better fitting of the stock replenishment parameters to a dynamic demand.

1. Introduction

The automotive industry is, since its inception, one of the most competitive industries in the business world. With the fast development of automotive industry, the offer of automobile models with different power train, aesthetics and auxiliary functions market has really large dimensions. The reliability of a brand is an important concept in order to raise and keep customers loyalty. The safety of drivers and passengers is a major driver of the brand construction, and periodic maintenance is required in order to prevent vehicle failures and accidents. Preventive and corrective maintenance is behind the need for inventory management of spare parts (Ghorbel, 2014). Spare parts have to be available to guarantee an adequate level of service (Botter et al., 2000) and its inventory control is challenging when spare parts of many systems change over time (Vaisakh, 2013). Since spare parts tend to vary and have different costs, service requirements and demand patterns (Boylan et al., 2007) a negative impact on the indirect cost of the services is expected if a high quantity of spares is kept available for immediate use, meaning that a trade-off has to be made.

Demand and inventory management for spare parts is a complex subject as a result of the high number of parts managed (Cohen et al., 2006), the presence of intermittent or lumpy demand patterns (Boylan et al., 2010), the high responsiveness required due to downtime costs incurred by users (Murthy et al., 2004), the risk of stock obsolescence (Cohen et al., 2006).

While automotive industry is increasing its range of options, spare parts management gathers more diversity and complexity. Thus arises the need to practice good and efficient management of spares, taking into consideration objectives of low costs of stock together with high service levels. Having reasonable inventory is critical for improving the continuous efficiency for spare parts and decreasing the maintenance budget. In reality, on automotive maintenance companies, spares replenishment patterns do not always comply their actual demand. Thus, many unnecessary spare parts are piled in the warehouse, even though some demanded spares are in short. The application and use of spare part budget is not done according to effective guidelines to deal with demand uncertainty nature.

In a recent review, Ghorbel et al. (2014) present a classification of replenishment policies of spare parts inventory into static and dynamic groups. The former considers that replenishment parameters are constant in a specific period, while the later considers that they vary over time as far as they need to accommodate an uncertain environment that reflects the industrial reality. Rego et al. (2011) highlight that even the “reactive” models, which do not directly use demand forecasting to decide how much to order each time, have the need of medium-term demand forecasts for the definition of their parameters. In the majority of the surveyed

studies the demand forecasting model is selected taking into account the classification of parts (Rego et al. 2011). Also Huiskonen (2001) and Boylan et al. (2008b) consider that items classification has a crucial role in inventory management systems essentially because it can determine the adequate level of managerial attention, allow the choice of demand forecasting and inventory control methods, and establish different performance goals at the inventory turnover and service levels among the different classes of items.

Wang (2012) proposed a cost model to optimize the control the stock of spare parts in conjunction with preventive maintenance intervals of plant equipment. The model considers the random nature of plant failures. The main objective of this model is to find optimal solutions for the periodic review inventory policy together with preventive intervals, on a finite time horizon based on a dynamic programming. A dynamic inventory control model based on forecasting was developed by Babai et al. (2009). The results obtained using the dynamic procedure show similarity at service levels and superiority in storage costs when compared with a traditional static model.

This paper aims to present an operational model developed to improve the management of spare parts at automotive maintenance providers, in view of objectives of low costs of stock together with high service levels. To deal with significant differences among the demand patterns, the spares are classified based on measures of demand variability, allowing the selection of appropriated forecast methods to estimate the next period demand. It is then proposed to manage the replenishment of stock based on a dynamic reorder point and order quantity calculated considering at each time period the average estimated demand and measures of error of past estimates.

2. Methodology

The collection of data related with the orders registrations of maintenance components took place at an automotive maintenance provider for a particular vehicle model. The data comprise 1485 different spare references and more than 21,000 order demands during the year of 2014. Taking into account the extent of the analysis and time constraints associated with the study, the collected data was delimited. To select an appropriated sample of spare references a classification methodology was defined taking into account the annual demand and the unit costs of the components. By using an “ABC Classification”, the data was divided into three categories of importance according to a value based criterion reflecting the impact in the workshop’s annual turnover. The top 30 components considered as the most important, were selected as those to analyse.

3. Dynamic model for stock replenishment

The first step of the proposed methodology is responsible for the identification of the demand patterns of each component based on some previous studies performed by authors as Croston (1972), Syntetos (2001, 2005) and Boylan et al. (2008). In fact, the different spares have quite different demand patterns: some showing regular and others presenting a quite irregular demand profile with several periods with a null demand. So, the concept of “Irregular Demand” introduced by Croston (1972) and the classification proposed by Syntetos (2001) as regards the adequacy of demand forecasting methods is considered. Syntetos classifies the demand patterns into four categories: (1) Intermittent, when some periods have zero demand, but when demand occurs it shows smooth values; (2) Erratic, when demand is highly variable, but the demand time intervals are mostly constant; (3) Smooth, when demand size and time intervals are both regular; and (4) Lumpy, when demand size is highly variable and the interval between demands is large, showing several time periods with zero demand. He also identified among a set of forecast methods the most performant ones for each demand pattern, eliciting the thresholds of the classes in regard to the Average Inter-Demand Interval (*ADI*) and Square Coefficient of Variation (CV^2) of no null demand (Figure 1). Original Croston’s Method (CO) and Single Exponential Smoothing (SES) were identified as the most appropriated forecasting methods considering the mean square error for performance evaluation (Syntetos, 2001). Later, Boylan et al. (2008a) proposed the Syntetos-Boylan Approximation Method (SBA), in fact a modification of the CO method, claiming its better performance for irregular demands.

This classification was applied to 30 selected components. The results obtained (Figure 2) show that 30% of the 30 components have an Intermittent demand profile, and 17%, 27%, and 23% have a Smooth, Lumpy and Erratic demand respectively. The SBA was then used in parallel with the CO method for testing

their demand forecasting performance. The results for both methods are compared as regards error measures allowing the selection of the most appropriate for each component.

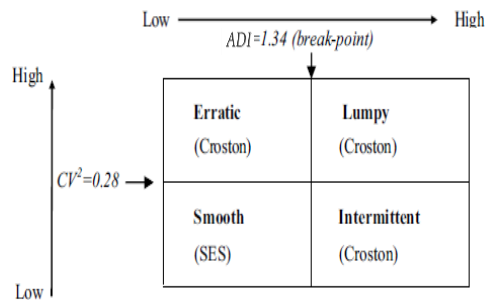


Figure 1 - Categorization of Demand Patterns by forecast accuracy of SES and CO forecasting methods (Boylan et al., 2008a).

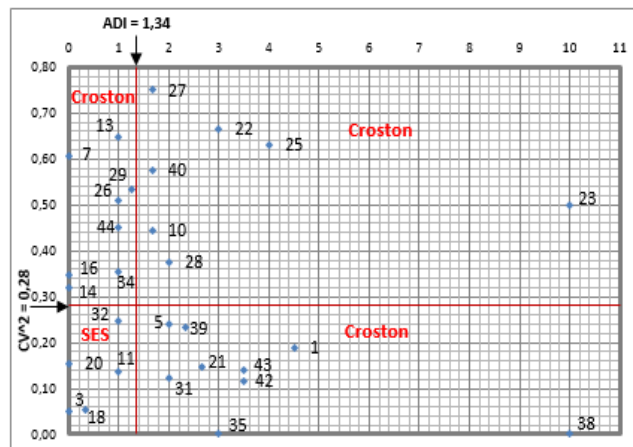


Figure 2 – Classification of components regarding their demand pattern.

Figure 3 presents the demand for four components, each associated to a different demand category. After having associated the appropriate forecasting method with each component (by fitting the methods coefficients towards a minimal error measure), the forecasts (F_t) can be calculated.

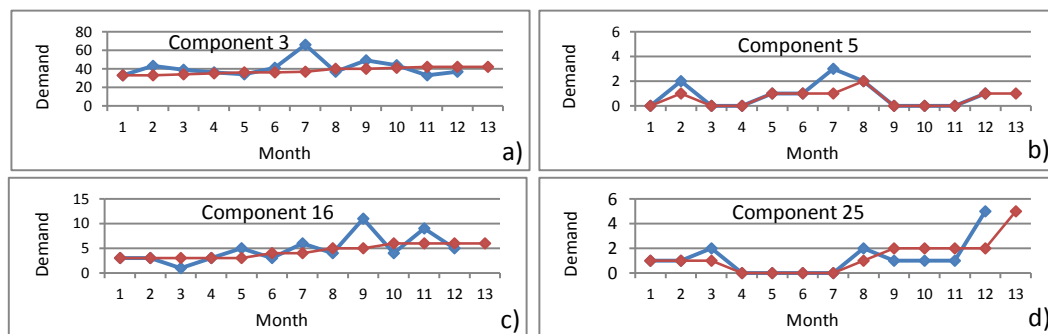


Figure 3 – Demand (blue line) versus Forecast (red line). a) Smooth behavior (SES); b) Intermittent behaviour (CO); c) Erratic behaviour (SBA); d) Lumpy behaviour (CO).

Being the most common in automotive maintenance providers, the Reorder Point replenishment policy was used for the spares. It considers that the stock level is kept under continuous observation, and new orders are made immediately when it falls below a reference level, the Reorder Point (R), defined by:

$$R = \bar{D} \cdot LT + SS \quad (1)$$

Where $\bar{D}$ is the average demand (per day), LT the order lead time (days) and SS the safety stock to accommodate the unpredictable fluctuations in demand during lead time, defined based on a intended service level, SL . The formulation for the security stock, assuming a normal distribution of demand, is:

$$SS = Z_{\alpha} \cdot \sigma_{LT} \quad (2)$$

In which Z_{α} is the z-score, corresponding to a certain service level, and σ_{LT} represents the demand standard deviation during replenishment lead time.

In the developed dynamic model the Reorder Point is automatically adjusted considering the forecasted demand for the following month and the updating over time of σ_{LT} based on past demand. The root square error (RSE) of past forecasts is used to estimate σ_{LT} :

$$\sigma_t = \sqrt{\frac{\sum_{i=1}^t (RSE_i - \overline{RSE})^2}{t}} \quad \text{and} \quad \sigma_{LT} = \sqrt{\frac{LT}{wd}} \cdot \sigma_t \quad (3)$$

Where t refers to the forecast period (month) and wd is the average number of working days per month.

Also the dimensioning of Q takes into account the demand forecast of each month (F_t) and its forecast deviation error (σ_t). In the formulation (4), MNO represents the “mean number of orders” per month:

$$Q = \frac{F_t}{MNO} + \frac{\sigma_t}{\sqrt{MNO}} \quad (4)$$

4. Model's Application

Figure 4 shows some results of Reorder Point dynamic model applied to the four chosen components. The Reorder Point is recalculated and adjusted to the expected demand for the next month. Notice that the first period (month 1) is considered as the initialization period and the last period (month 13) is now ready to receive the input of the real demand, enabling the model to evolve for future demand situations.

SL = 95%		Component 3 – Smooth							
z = 1.64		$\alpha = 0.1$		SES					
Month (t)	D_t	F_{t+1}	RSE_t	σ_t	σ_{LT}	SS	R	$R_{(t+1)}$	
Lead Time	2								
1	33	33	-	-	-	-	-	-	
2	43	33	10.0	7.07	2.14	4	7		
3	39	34	6.0	5.03	1.51	3	7		
4	36	35	2.0	4.43	1.34	3	7		
5	34	36	1.0	4.15	1.24	3	7		
6	41	36	5.0	3.74	1.13	2	6		
7	66	37	30.0	10.40	3.14	6	10		
8	37	40	0.0	10.01	3.01	5	9		
9	49	40	9.0	9.39	2.83	5	9		
10	44	41	4.0	8.91	2.69	5	9		
11	33	42	8.0	8.46	2.55	5	9		
12	37	42	5.0	8.08	2.43	4	8		
13	-	42	-	-	-	-	-	-	

a)

SL = 95%		Component 5 - Intermittent							
z = 1.64		$\alpha = 0.25$		CO					
Month (t)	D_t	F_{t+1}	RSE_t	σ_t	σ_{LT}	SS	R	$R_{(t+1)}$	
Lead Time	2								
1	0	0	-	-	-	-	-	-	
2	2	1	2.0	1.41	0.42	1	2		
3	0	0	1.0	1.00	0.30	1	1		
4	0	0	0.0	0.96	0.28	1	1		
5	1	1	1.0	0.84	0.25	1	2		
6	1	1	0.0	0.82	0.24	1	2		
7	3	1	2.0	0.90	0.27	1	2		
8	2	2	1.0	0.83	0.25	1	2		
9	0	0	2.0	0.87	0.25	1	1		
10	0	0	0.0	0.88	0.27	1	1		
11	0	0	0.0	0.87	0.27	1	1		
12	1	1	1.0	0.83	0.25	1	2		
13	-	1	-	-	-	-	-	-	

b)

SL = 95%		Component 16 - Erratic							
z = 1.64		$\alpha = 0.3$		SBA					
Month (t)	D_t	F_{t+1}	RSE_t	σ_t	σ_{LT}	SS	R	$R_{(t+1)}$	
Lead Time	2								
1	3	3	-	-	-	-	-	-	
2	3	3	0.0	0.00	0.00	0	1		
3	1	3	2.0	1.15	0.35	1	2		
4	3	3	0.0	1.00	0.30	1	2		
5	5	3	2.0	1.10	0.33	1	2		
6	3	4	0.0	1.03	0.31	1	2		
7	6	4	2.0	1.07	0.33	1	2		
8	4	5	0.0	1.04	0.31	1	2		
9	11	5	6.0	2.00	0.61	1	2		
10	4	6	1.0	1.89	0.57	1	2		
11	9	6	3.0	1.86	0.57	1	2		
12	5	6	1.0	1.78	0.54	1	2		
13	-	6	-	-	-	-	-	-	

c)

SL = 95%		Component 25 - Lumpy							
z = 1.64		$\alpha = 0.85$		CO					
Month (t)	D_t	F_{t+1}	RSE_t	σ_t	σ_{LT}	SS	R	$R_{(t+1)}$	
Lead Time	2								
1	1	1	-	-	-	-	-	-	
2	1	1	0.0	0.00	0.00	0	1		
3	2	1	1.0	0.58	0.17	1	2		
4	0	0	1.0	0.58	0.17	1	1		
5	0	0	0.0	0.55	0.17	1	1		
6	0	0	0.0	0.52	0.16	1	1		
7	0	0	0.0	0.49	0.14	1	1		
8	2	1	2.0	0.76	0.23	1	2		
9	1	2	0.0	0.73	0.21	1	2		
10	1	2	1.0	0.71	0.21	1	2		
11	1	2	1.0	0.69	0.21	1	2		
12	5	2	3.0	0.97	0.30	1	2		
13	-	5	-	-	-	-	-	-	

d)

Figure 4 – Reorder Point Dynamic Model. a) Smooth, b) Intermittent, c) Erratic, d) Lumpy.

The model's application considers a pre-defined maximum stock, which reflects the warehouse available space for each component. Each time a new demand is registered, the model calculates the new available quantity in stock, considering also the on-going orders. The available stock is then compared to the reorder point and a message is shown: in case it is higher than R the message “ok” is shown and no order is expected; in case it is lower a “Verification” function tells that an order has to be made showing the message “Order”.

The model was applied using the data collected in the maintenance provider and the results were compared against the current situation in which both the reorder point and the order quantity are static and assume the values used by the company (baseline static model). Two cases were analysed. Firstly, (R dynamic model) the model was run considering that the reorder point is dynamic, but the ordered quantity (Q) is the one necessary to replace the level of stock to its maximum. In fact this is the order quantity

normally used by the maintenance provider, so, somehow, Q behaves as a static parameter in this first model's application in the sense it is defined by the difference between maximum stock and the reorder point. Secondly, changes were made, and also order quantity was allowed to have a dynamic behaviour (R and Q dynamic model), adjusting to each verified demand.

The results obtained were discussed focusing on some criteria and compared to the baseline situation, namely as regards to the number of orders made, number of stock-outs, number of inoperative vehicles and mean stock level.

5. Model Application's Results

Taking into account the previous defined criteria, the results for each model are illustrated in Figure 5. The results are compared, starting with the analysis of the baseline static model, followed by the reorder point and reorder point and order quantity dynamic models.

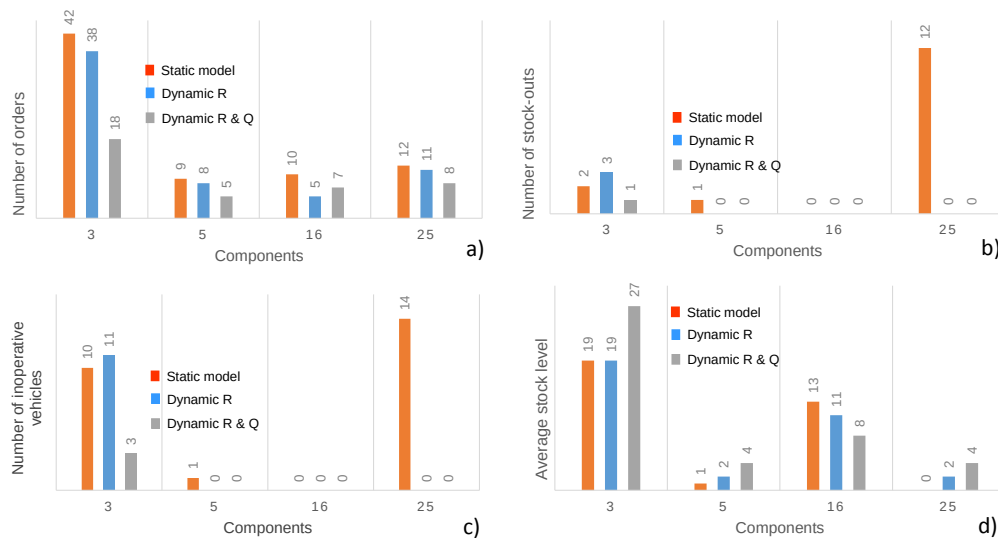


Figure 5 – Model Application's Results. a) Orders; b) Stock-outs; c) Inoperative vehicles; d) Mean stock. (it was assumed a mean number of orders per month of 1, following the current company practice for most components)

A reduction on the number of orders made was verified with the application of the dynamic models and, except in component 16, the lower values are given by the R and Q dynamic model.

The results for the number of stock-outs vary not only from model to model, but also from component to component. The component 3 registers a slightly raise in the number of stock-outs when comparing the baseline model with the R dynamic model, however the R and Q dynamic model performs always as good as or better than all the others. Component 25, which have a lumpy demand pattern, presents a large number of stock-outs for the baseline model, though no stock-outs are shown when the dynamic models are implemented.

The results for inoperative vehicles are derived from the stock-outs. A vehicle is considered as inoperative every time it awaits the availability of a given component in stock after a stock-out situation is registered. It is indeed the performance metric that mostly concerns the maintenance provider as it truly affects clients' satisfaction and induces significant direct costs in delivering substitution vehicles. The main concerns are for component 3 and component 25, which in the baseline have a significant impact in delaying vehicles. In both the dynamic approaches reduce or even eliminate that problem.

The mean stock level, in most cases, shows opposite results to the others before. In fact, the reduction in the number of stock-outs and inoperative vehicles of component 25 was achieved with a slightly absolute increase of its average stock. Something similar happens with components 3 and 5. However, the good performance of component 16, in the previous criteria was possible to be kept with a lower level of stock when dynamic models are used.

6. Conclusions

This paper led to the proposal of an integrated practice of forecasting demand for spare parts and dynamic regulation of replenishment parameters of a reorder point policy for stock management, which can easily be implemented into the information currently used by automobile maintenance providers. The analysis presented gives an overview based on a set of components and a few evaluation criteria. In the stock management theme, the ideal situation is clearly a situation where the best compromise emerges based on a combination of different criteria. In fact, to make more frequently regular orders with smaller quantities can be an objective to pursue, particularly if ordering costs can be kept under control. The stock control based on the dynamic models allow the reduction on the number of stock outs, leading to a more responsive service with lower mean stock levels. The developed methodologies may be applicable and generalized to other components and sectors. Results achieved allow believing that the dynamic models are a sounder solution for responsive maintenance services. On-going work involves the validation with a statistically significant sample of components and the analysis of the better forecasting period.

References

- Babai, M.Z., Syntetos, A.A., Dallery, Y., Nikolopoulos, K. (2009) Dynamic re-order point inventory control with lead-time uncertainty: analysis and empirical investigation. *International Journal of Production Research* 47 (9), pp. 2461-2483.
- Botter, R., Fortuin, L. (2000) Stocking strategy for service parts: a case study. *International Journal of Operation and Production Management* 20, pp. 656-674.
- Boylan, J.E., Syntetos, A.A. (2010) Spare parts management: a review of forecasting research and extensions, *IMA Journal of Management Mathematics* 21, pp. 227-237.
- Boylan, J.E., Syntetos, A.A. (2008a) Forecasting for Inventory Management of Service Parts. In: *Complex System Maintenance Handbook*. Eds: Kobbacy, K.A.H., Prabhakar Murthy, D.N., pp. 479-506. London: Springer.
- Boylan, J.E., Syntetos, A.A., Karakostas, G.C. (2008b) Classification for forecasting and stock control: a case study. *Journal of the Operational Research Society* 59, pp. 473-481.
- Boylan, J.E., Syntetos, A.A. (2007) The accuracy of a modified Croston procedure. *International Journal of Production Economics* 107, pp. 511-517.
- Cohen, M.A., Agrawal, N. (2006) Winning in the Aftermarket. *Harvard Business Review*, v.84, pp. 129-138.
- Croston, J. D. (1972) Forecasting and Stock Control for Intermittent Demand. *Operational Research Quarterly* 23(3), pp. 289-303.
- Ghorbel, N., Addouche, S., El-Mhamedi, A., Boujelbene, Y. (2014) Replenishment policies in static and dynamic spare parts inventory control: A survey. *International Journal of Engineering Development and Research*. 2 (4), pp. 3636-3655.
- Huiskonen, J. (2001) Maintenance spare parts logistics: Special characteristics and strategic choices", *International Journal of Production Economics* 71, pp. 125-133.
- Murthy, D.N.P., Solem, O., Roren, T. (2004) Product warranty logistics: issues and challenges. *European Journal of Operations Research* 156, pp. 110-126.
- Rego, J.R., de Mesquita, M.A. (2011) Spare parts inventory control: a literature review, *Produção* 21 (4), pp. 656-666, doi: 10.1590/S0103-65132011005000002.
- Syntetos, A.A. (2001) *Forecasting of Intermittent Demand*, Unpublished Phd Thesis, UK: Buckinghamshire Chilterns University College, Brunel University.
- Syntetos, A.A., Boylan, J.E. (2005) The accuracy of intermittent demand estimates. *International Journal of Forecasting* 21.
- Syntetos, A.A., Boylan, J.E., Disney, S.M. (2009) Forecasting for inventory planning: a 50-year review. *Journal of the Operational Research Society* 60, pp. S149-S160.
- Vaisakh P.S., Dileepal J., Unni, V.N. (2013) Inventory Management of Spare Parts by Combined FSN and VED (CFSNVED) Analysis, *International Journal of Engineering and Innovative Technology* 2 (7).
- Wang, W. (2012) A stochastic model for joint spare parts inventory and planned maintenance optimization. *European Journal of Operational Research* 216, pp.127-139.

Optimising life extension and maintenance policy decisions for safety critical elements

Isaac Animah, Mahmood Shafiee*, Nigel Simms

Cranfield University, College Road, Bedfordshire MK43 0AL, United Kingdom

* Corresponding author, email: m.shafiee@cranfield.ac.uk

Abstract

Extending the service life of safety critical elements has been of great interest to asset managers in the UK. In this paper, we present an optimisation model to determine the optimal length of life extension and PM strategy such that the expected total cost of safety critical elements during the extended life phase is minimized, assuming that the failure rate of the system's lifetime is bathtub-shaped.

Keywords. Imperfect maintenance, life expectancy, life extension, maintenance optimization, mean remaining lifetime, reliability, safety critical asset.

Introduction

The growing demand of energy worldwide, lack of capital investment by the private sector, and highly restrictive regulations prevent industries from constructing new plants. For these reasons, over the past decade, the extension of the lifetime of safety critical assets has increasingly attracted attention of many scholars and practitioners in the field of reliability and asset management. The life extension as an end-of-life strategy places a burden on asset managers to improve the condition of assets such that a continuous, safe and reliable operation beyond original design life is guaranteed.

Most of the assets operating in safety critical industries such as offshore oil and gas, nuclear power, petrochemical, renewable energy, rail transport, aviation, shipping, and electricity distribution and transmission suffer from time-dependent degradation failures that occur during the last stages of the lifecycle. These failures cause the whole system to operate in a degraded mode, resulting in reduced efficiency and safety. This implies that appropriate actions must be employed at the life extension phase of assets to reduce the number of failures, thereby increasing system's availability and reliability. Ramírez and Utne (2015) suggested that one of the effective ways to reduce failure intensities of safety critical elements when they reach the end of service life is to increase the frequency of maintenance actions. Feinstein and Morris (2010) indicated that the integration of maintenance strategy with asset management policy could play a critical role in keeping the assets in good working condition during life extension period. Therefore, it is crucial to develop an appropriate maintenance policy at the end of assets' useful lifespan in order to reduce failure intensity, increase availability and improve reliability of safety critical elements.

According to Shafiee (2015), the maintenance actions for managing or repair of industrial assets can be classified as either corrective or preventive. A corrective maintenance (CM) is an unplanned action that is performed to restore the equipment when they fail or break down, while a preventive maintenance (PM) is a pre-planned action performed to improve the functional performance of equipment before a failure occurs. The CM/PM actions can either restore a system's operating condition to "as good as new (AGAN)" (perfect repair) or leave the system in the state it was in prior to the failure, i.e., "as bad as old (ABAO)" (minimal repair). However, the reality on the field is that maintenance actions may sometimes restore the asset's operating state to somewhere between AGAN and ABAO. Such maintenance actions are called *imperfect maintenance/repair*.

In recent years, imperfect CM and PM models based on reduction of failure intensity or effective age have received much attention in reliability and maintenance literature. Doyen and Gaudoin (2004) proposed a generic imperfect maintenance model in which the repair effect is represented by a reduction of the

system's age. Cassady *et al.* (2005) evaluated the impact of imperfect maintenance on the long-run availability of repairable systems using the virtual age process. Dijoux (2009) developed a virtual age model for repairable systems whose failure rate follows a bathtub curve. The main characteristics of the model is that it incorporates the effect of maintenance on extension of the system's remaining useful lifetime (RUL). Sheu *et al.* (2010) presented a generic imperfect PM model based on virtual age approach for systems subjected to random shocks. Tsai *et al.* (2011) developed an age-reduction model for repairable systems that undergo imperfect maintenance actions at pre-scheduled times and minimal repair actions at failures. Shafiee *et al.* (2013) developed a virtual age model to determine optimal length of burn-in time, optimal number of PM actions and the associated degree of repair such that warranty servicing cost is minimized. Further application of the imperfect maintenance models to the determination of optimal maintenance interval could be found in Pham and Wang (2006), Bartholomew-Biggs *et al.* (2009) and Liu and Huang (2010).

Increasing the number of PM activities over the life extension phase of operation will improve the reliability of the system by reducing the deterioration rate of safety critical elements but on the other side, it will also result in an increase in the costs. For this reason, determining the optimal frequency and degree/quality of repair actions during the extended period of operation is of great importance to asset managers across various industries. In spite of the vast literature concerning determination of PM intervals over original design lifetime, a survey of the literature shows that there are very few studies in relation to decision making on the duration of lifetime to be prolonged and the PM policy to be adopted.

Method

Figure 1 illustrates the processes involved in the development of the life extension and maintenance strategy decision-making model. As shown, the decision model requires some inputs such as historical data including maintenance records, failure records, system information, cost data, etc. The framework comprises two main steps: (i) development of the imperfect maintenance model and construction of the cost model and (ii) formulation of the optimization problem and description of the solution method.

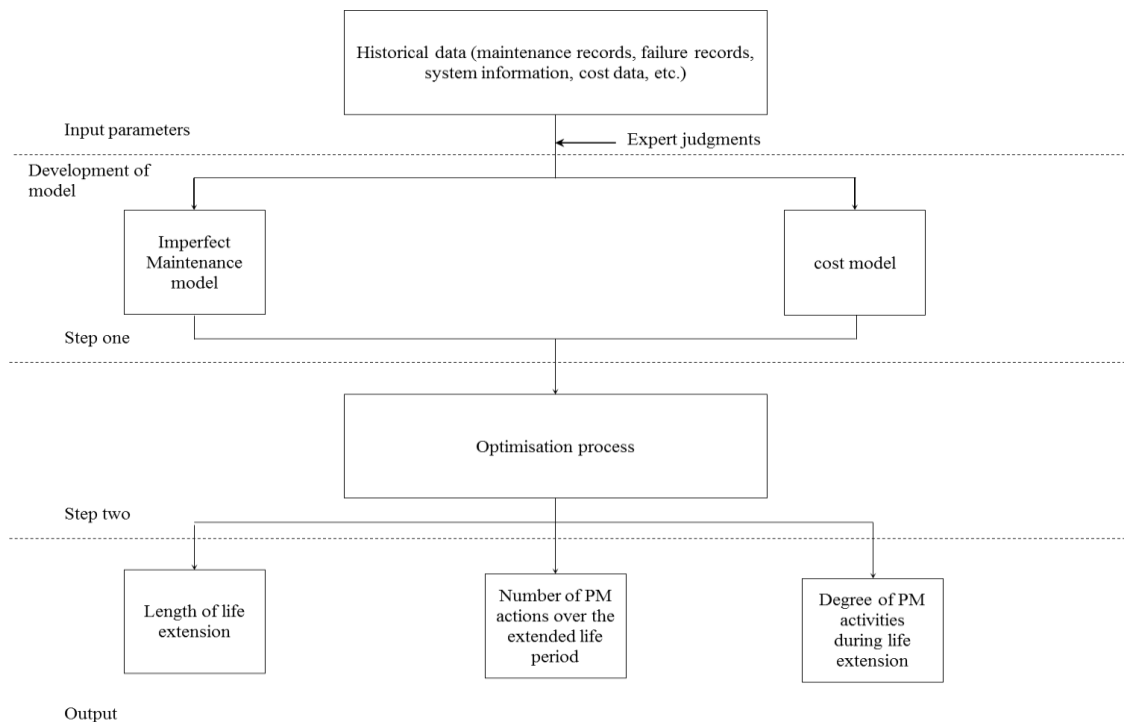


Figure 1. Model development

Conclusion

The model is based on the virtual age concept and a novel cost model representing the relationship between age reduction and the cost of PM actions is proposed. Imperfect PM actions are carried out at discrete time instants and their effects are represented by reduction of chronological age of the asset. The cost of imperfect PM actions is considered as an increasing function of the age reducing factor. The age-reduction PM scheme during the extended life of operation is illustrated in Figure 2. A non-linear mathematical model is formulated to solve the addressed problem and the necessary and sufficient conditions for the existence of an optimal solution when the time-to-failure of asset follows Weibull distribution are obtained. Finally, a numerical study is provided to illustrate the results and a sensitivity analysis is conducted to evaluate the effect of model parameters on the optimal solution.

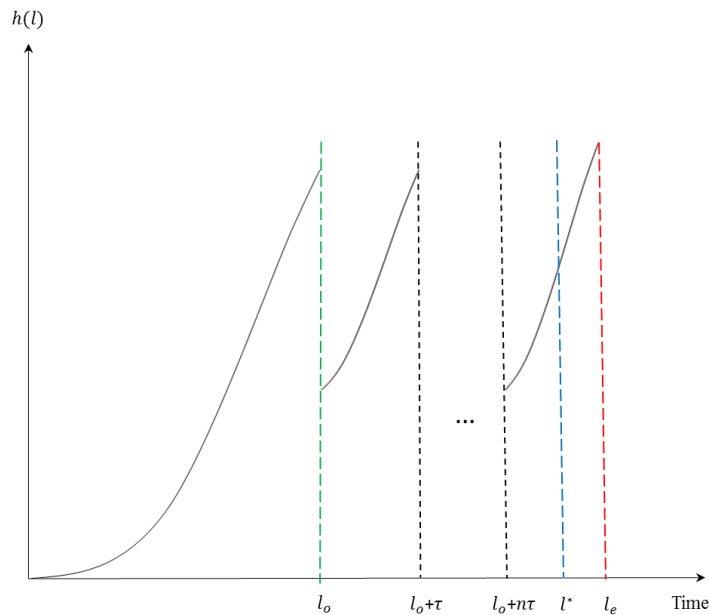


Figure 2. An age-reduction preventive maintenance scheme.

References

- Bartholomew-Biggs, M., Zuo, M.J. & Li, X., 2009. Modelling and optimizing sequential imperfect preventive maintenance. *Reliability Engineering and System Safety*, 94, pp.53–62.
- Cassady, C.R., Iyob, I.M., Schneider, K. & Pohl, E.A., 2005. A generic model of equipment availability under imperfect maintenance. *IEEE Transactions on Reliability*, 54(4), pp.564–571.
- Dijoux, Y., 2009. A virtual age model based on a bathtub shaped initial intensity. *Reliability Engineering & System Safety*, 94(5), pp.982–989.
- Doyen, L. & Gaudoin, O., 2004. Classes of imperfect repair models based on reduction of failure intensity or virtual age. *Reliability Engineering and System Safety*, 84(1), pp.45–56.
- Feinstein, C.D. & Morris, P. a, 2010. The role of uncertainty in asset management. *IEEE PESTransmission and Distribution Conference and Exposition*, pp.1–6.
- Liu, Y. & Huang, H.-Z., 2010. Optimal Replacement Policy for Multi-State System Under Imperfect Maintenance. *IEEE Transactions on Reliability*, 59(3), pp.483–495..
- Pham, H. & Wang, H., 2006. A Quasi-renewal Process and Its Applications. In *Reliability and Optimal Maintenance*. pp. 51–90.
- Ramírez, P. a. P. & Utne, I.B., 2015. Use of dynamic Bayesian networks for life extension assessment of ageing systems. *Reliability Engineering & System Safety*, 133, pp.119–136..
- Shafiee, M., Finkelstein, M. & Zuo, M.J., 2013. Optimal burn-in and preventive maintenance warranty strategies with time-dependent maintenance costs. *IIE Transactions*, 45(9), pp.1024–1033
- Shafiee, M., 2015. Maintenance strategy selection problem: An MCDM overview. *Journal of Quality in Maintenance Engineering*, 21(4), pp.378–402.
- Sheu, S.-H., Chang, C.-C. & Chen, Y.-L., 2010. A generalized periodic preventive maintenance model with virtual age for a system subjected to shocks. *Communications in Statistics - Theory and Methods*, 39(13), pp.2379–2393.
- Tsai, T-R., Liu, P-H. & Lio, Y.L., 2011. Optimal maintenance time for imperfect maintenance actions on repairable product. *Computers & Industrial Engineering*, 60(4), pp.744–749.

,

Modelling on Joint Optimization for Two-Dimensional Age Replacement Interval and Spare Parts Provisioning

Bai Yongsheng, Hu Qiwei, Zhao Jianmin

Department of Management Engineering, Mechanical Engineering College, No.97, Heping West Road, Shijiazhuang, China, baiys820708@sina.com

Abstract. To achieve high availability with minimal maintenance and support cost, the relevant decision variables, such as PM interval, maximum spare part inventory level etc. should be considered jointly. In this paper, for items whose degradation and maintenance are measured both by calendar and usage time, the maintenance strategy of two-dimensional age replacement is introduced, and the corresponding spare parts provisioning process is analyzed. On the basis of spare parts inventory cost analysis, the joint optimization model for two-dimensional age replacement interval and spare parts provisioning is established, which can minimize the total cost per unit time. Finally, a case study is given to demonstrate the model's applicability and validity.

Key Words: two-dimensional maintenance; age replacement; spare parts provisioning; joint optimization; cost model

1. Introduction

As is known, the provisioning of spare parts plays an important role in engineering asset management, and there are a lot of related literature, dealing with demand forecast, supply strategy, level of repair, stock allocation etc (Basten R.J.I.2012; Deniz Caglar 2004; Kennedy W. J. 2002; Seongmin Moon 2012) . However, the researches above rarely focused on the decision-making under preventive maintenance; while most preventive maintenance models neglect whether the spare parts inventory is adequate, and assume that the replacement items are drawn from an infinite stock. Obviously, it is lack of scientificity and pertinency, which not only leads to the waste of maintenance and stock cost, but decreases the system readiness and availability.

So, the study on joint optimization of preventive maintenance and spare parts provisioning is necessary. Karin etc. (2009) integrated preventive maintenance frequency, spare part inventory levels and spare part repair capacity, and developed heuristics algorithm for k -out-of- N systems under condition-based maintenance and block replacement, respectively. Xu C. etc. (2011) studied the spare part demand for a single-unit system, and established a cost optimization model based on the comprehensive consideration of preventive maintenance policy and spare part inventory. Darko Louit etc. (2011) studied the condition-based spares ordering for critical components, and determined ordering decision depending on the remaining useful life obtained through the assessment of component age and condition indicators at every inspection time.

In real life, the failures of many items are time and usage dependent (Baik J. 2004; Jerry Lawless 1995; Jung M. 2007; Tom Chen 2002). For these items, the preventive maintenance tasks are implemented under so-called two-dimensional maintenance policy. A common example of such items is the automobile whose life is measured by both time in service and mileage, and its preventive maintenance usually applies an interval of six months or five thousand kilometers, whichever comes first. Recently, the two-dimensional maintenance mostly exists in warranty research field, used to determine the warranty period. The optimization studies on spare parts provisioning under two-dimensional preventive maintenance above are rare.

Hence in this paper, we consider joint optimization of maintenance interval and spare parts inventory under two-dimensional preventive replacement policy. The outline of the paper is as follows. The characteristics of spare parts demand and inventory consumption process for two-dimensional preventive maintenance is analyzed in Section 2. In Section 3, we present the total cost structure in maintenance and spare parts provisioning process, deduce the probability of maintenance events, and establish the joint optimization model for two-dimensional age replacement interval and spare parts provisioning. In Section 4,

a numerical example is given, which demonstrates the model's applicability and validity. Concluding remarks and suggested further research are presented in Section 5.

2. Analysis of item maintenance and spare parts provisioning process

2.1 Problem description

Generally, the implementation of maintenance activities needs spare parts, especially for periodic replacement tasks. Thus, the availability of spare parts is an important factor in maintenance efficiency and effectiveness (Martorell S. 2002; Tsai Y.T. 2004). As is known, there are clear relations between maintenance interval and spare parts provisioning (Hu Qiwei 2015; Wang Wenbin 2011). For example, a shorter preventive replacement interval leads to a higher spare parts demand from preventive maintenance, meanwhile a lower spare parts demand from corrective replacement, and vice versa.

Therefore, the joint optimization of spare parts inventory and maintenance policies is an effective way to reduce maintenance cost and increase system availability. As mentioned above, the maintenance timing of many items is determined both by calendar and usage time. Hence, this paper deals with the situation of two-dimensional maintenance interval (T_0, U_0) , that is, the item is preventively replaced according to both calendar limit T_0 and usage limit U_0 , whichever comes first. And the regular ordering policy (t_0, S) is adopted for spare parts inventory control, that is, the stock level is replenished to S with t_0 as ordering period. The objective is to establish a model considering replacement period and spare parts inventory-control simultaneously in infinite time horizon, which can minimize the total cost per unit time by optimizing (T_0, U_0) and (t_0, S) jointly.

2.2 Inventory consumption process analysis under two-dimensional age replacement

Under the maintenance policy of two-dimensional age replacement, the preventive replacement is determined both by calendar and usage time, so there are three situations for item replacement, that is, preventive replacement based on the calendar time T_0 , preventive replacement based on the usage time U_0 , and corrective replacement based on the stochastic failure time t , as illustrated in Figure 1. Obviously, the spare parts demand just derives from the same situations above.

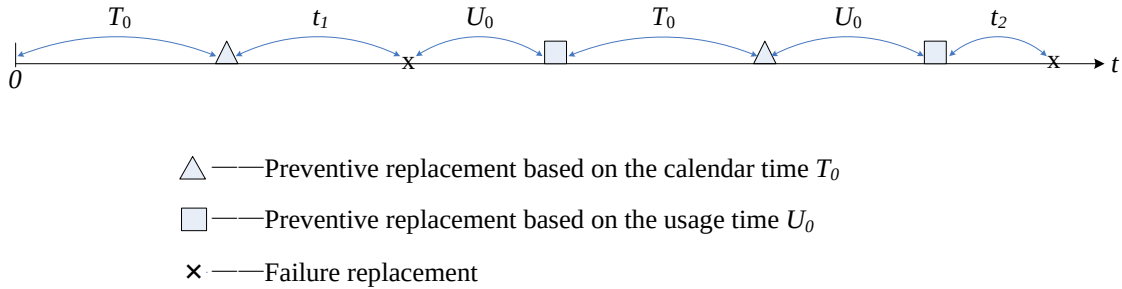


Figure 1. Illustration of two-dimensional age replacement policy

Recently, there are essentially three approaches to model failures for items with two-dimensional interval (Mahmood Shafiee 2011). In the first approach, the failure of items is modelled using a two-dimensional point process formulation with bivariate distribution function $F(t, u)$ (Corbu D. 2008; Kim H.G 2000; Pal S. 2003; Yang S.C 2001). In the second approach, the two measures--calendar and usage time--are combined to provide a single composite scale Z (e.g., $Z=at+bu$), by which the failure process is modelled (Duchesne T. 2000; Gertsbakh I.B 1998)¹. In the third approach, it's assumed that the item usage rate R ($R=U/T$) varies from customer to customer, but it is a constant for a given customer (Chukova S. 2006; Jack N. 2009; Manna D.K 2006; Rai B. 2005). Then, R can be treated as a random variable with given cumulative distribution function $F_R(r)$, $0 \leq r < \infty$. In this paper, the third approach is adopted.

Clearly, the spare parts provisioning process with inventory policy (t_0, S) has no more than the following two situations. In situation 1, the spare parts still have the rest when it reaches t_0 ; in situation 2, the spare parts are run out of before t_0 . However, under two-dimensional age replacement strategy, the interval of preventive replacement can be T_0 or U_0 . Thus, the inventory consumption process of either

situation mentioned above can further be divided into two cases, that is, case A: $r \leq r_0$, and case B: $r > r_0$ ($r_0 = \frac{U_0}{T_0}$). When the usage rate r is less than r_0 , the age replacement interval is T_0 ; on the contrary, the age replacement interval is $\frac{U_0}{r}$, as shown in Figure 2 and Figure 3.

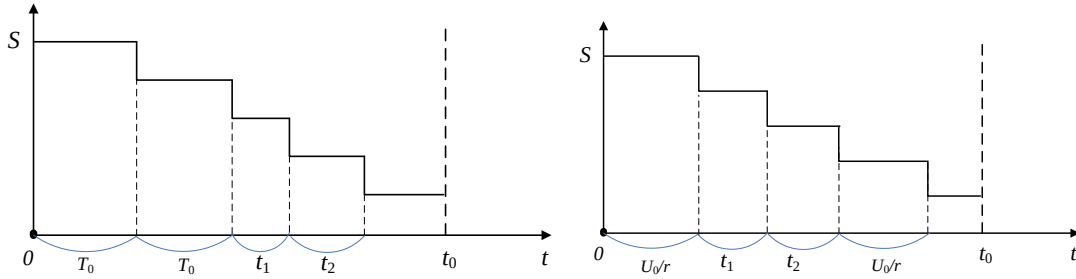


Figure 2-A. Inventory consumption process with sufficient stock ($r \leq r_0$)

Figure 2-B. Inventory consumption process with sufficient stock ($r > r_0$)

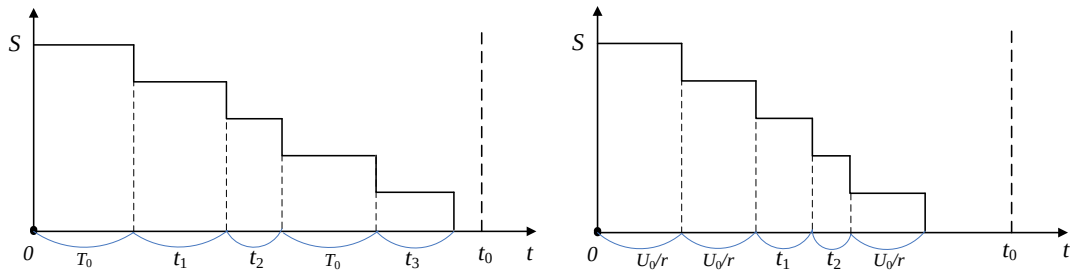


Figure 3-A. Inventory consumption process with inadequate stock ($r \leq r_0$)

Figure 3-B. Inventory consumption process with inadequate stock ($r > r_0$)

Through above analysis, the spare parts inventory consumption process under two-dimensional age replacement is derived, and the item maintenance and spare parts provisioning process can be mathematically modelled further.

3. Modelling total cost for item maintenance and spare parts provisioning

For the optimization of total cost in item maintenance and spare parts provisioning process, the cost structure is analyzed below, and the mathematical models are established in steps.

3.1 Assumptions and notations

- The run time of spare parts system is long enough to be treated as infinite time horizon.
- The ordering period is far greater than supply time, and the delay time for supply can be ignored.
- The item should be replaced by a new one at the time of either a failure or preventive replacement interval.
- T : calendar time of the item, e.g. day, week, month, and year etc.
- U : usage time of the item, e.g. mile, kilometer, operational etc.
- (T_0, U_0) : two-dimensional age replacement interval.
- t_0 : spare parts ordering period.
- S : spare parts maximum stock level.
- C_p : the average cost per age replacement.
- C_f : the average cost per failure replacement.
- C_s : the expected storage cost within an ordering period.
- C_o : the expected spare parts ordering cost within an ordering period.
- C_{los} : the expected loss due to stock shortage within an ordering period
- $F_R(r)$, $f_R(r)$: the cumulative distribution function and probability density function of the usage rate.

- r_l, r_u : the lower and upper bound of usage rate.
- $F(t; r), f(t; r)$: the cumulative distribution function and probability density function of item failure time given fixed usage rate.
- $EC(T_0, U_0, t_0, S)$: the expected total cost per unit time for spare parts system with two-dimensional maintenance interval (T_0, U_0) and inventory control policy (t_0, S) .

3.2 Maintenance and spare parts provisioning cost analysis

In infinite time horizon, each inventory replenishment can be seen as the renewal of spare parts system. According to the renewal theorem, the expected total cost per unit time for spare parts system in infinite time horizon can be expressed as the expected cost per unit time within an ordering period. Thus, combining all the possible cases of inventory consumption process, we have the expected total cost per unit time as:

$$\begin{aligned} EC(T_0, U_0, t_0, S) &= EC_1(T_0, U_0, t_0, S) + EC_2(T_0, U_0, t_0, S) \\ &= \int_{r_l}^{r_0} EC_{1A}(T_0, U_0, t_0, S; r) \cdot f_R(r) dr + \int_{r_0}^{r_u} EC_{1B}(T_0, U_0, t_0, S; r) \cdot f_R(r) dr \\ &\quad + \int_{r_l}^{r_0} EC_{2A}(T_0, U_0, t_0, S; r) \cdot f_R(r) dr + \int_{r_0}^{r_u} EC_{2B}(T_0, U_0, t_0, S; r) \cdot f_R(r) dr \end{aligned} \quad (1)$$

Where $EC_{1A}(T_0, U_0, t_0, S; r)$ is the expected total cost per unit time with fixed r for case A in situation 1 as mentioned in Section 2.2, where the two-dimensional maintenance interval (T_0, U_0) and inventory policy (t_0, S) is employed. And the rest notations in equation 1 have similar meanings.

Corresponding to the inventory consumption process, the total cost in maintenance and spare parts provisioning can be divided into two situations. Specifically, in situation 1, when the spare parts still have the rest at t_0 , the costs include age replacement cost C_p , failure replacement cost C_f , storage cost C_s , and ordering cost C_o ; in situation 2, for the spare parts are run out of before t_0 , the costs also include stock shortage loss C_{los} , in addition to the above costs in situation 1.

The main purpose for modelling is to consider replacement period and spare parts inventory-control simultaneously, analyze the probabilities of various event occurred in maintenance and spare parts provisioning process, and express the expected total cost per unit time with the above costs including C_p , C_o , C_{los} , and so on.

3.3 Modelling event probability and cost

The probabilities of events and cost in maintenance and spare parts provisioning process is analyzed and expressed in detail, and the joint optimization model for two-dimensional age replacement interval and spare parts provisioning is established below.

3.3.1 Analysis of life distribution for spare parts system under two-dimensional age replacement

According to the analysis in Section 2.2, it's known that the spare parts provisioning process has two situations, that is adequate inventory and insufficient inventory. When the total run time for spare parts system T_S is longer than the spare parts ordering period t_0 , situation 1 happens; on the contrary, if T_S is shorter than t_0 , then situation 2 happens. Obviously, the occurrence of events above is stochastic, so the life distribution for spare parts system needs to be deduced under two-dimensional age replacement strategy.

For usage rate approach is adopted to deal with two-dimensional replacement in this paper, if $T_b(T_r)$

denotes the time to first failure under usage rate $r_b(r)$, we have $\frac{T_r}{T_b} = \left(\frac{r_b}{r}\right)^\gamma$ with $\gamma \geq 1$. If the cumulative distribution function for T_b is given by $F(t; \alpha, \beta_b)$, where α and β_b is the shape parameter and scale

parameter respectively, the cumulative distribution function for T_r is the same as that for T_b but with scale

parameter given by $\beta(r) = \beta_b \cdot \left(\frac{r_b}{r}\right)^\gamma$. Hence, the cumulative distribution function for T_r is

$F(t; \alpha, \beta(r)) = F\left(t; \alpha, \beta_b \cdot \left(\frac{r_b}{r}\right)^\gamma\right)$. To express conveniently, it is abbreviated as $F(t; r)$. Similarly, the

corresponding probability density function $f(t; r)$ and failure rate function $\lambda(t; r)$ can be got.

Thus the cumulative distribution function for spare parts system life T_S can be derived as follows:

Let v_i denotes the life that a spare part can reach, and t_i denotes its actual life span, then the life of spare parts system T_S is the sum of $S+1$ items life span, that is

$$T_S = \sum_{i=0}^S t_i = \begin{cases} \sum_{i=0}^S \min\{v_i, T_0\}, & r \leq r_0 \\ \sum_{i=0}^S \min\left\{v_i, \frac{U_0}{r}\right\}, & r > r_0 \end{cases} \quad (2)$$

Transformed into the total probability form, the cumulative distribution function for T_S can be expressed as

$$\begin{aligned} G(t) &= P\{T_S < t\} \\ &= P\{T_S < t | n=0\} \cdot P(n=0) \\ &\quad + P\{T_S < t | n=1\} \cdot P(n=1) \\ &\quad + \dots \\ &\quad + P\{T_S < t | n=S\} \cdot P(n=S) \end{aligned} \quad (3)$$

where n denotes the times of age replacement.

For a fixed n , it's clear that

$$P(n=i) = \begin{cases} C_S^i \cdot R^i(T_0; r) \cdot F^{S-i}(T_0; r), & r \leq r_0 \\ C_S^i \cdot R^i\left(\frac{U_0}{r}; r\right) \cdot F^{S-i}\left(\frac{U_0}{r}; r\right), & r > r_0 \end{cases} \quad (4)$$

So the key problem is the expression of $P\{T_S < t | n\}$. Because there are n spare parts with actual life as (T_0, U_0) , the rest $S+1-n$ items must be replaced due to stochastic failure, and their total life is $t - n \cdot T_0$ or $t - n \cdot \frac{U_0}{r}$. Among $S+1-n$ items, the former $S-n$ of them fail before (T_0, U_0) , and the last runs until natural failure with no constraint. Hence, let X denotes the actual life of the former $S-n$ spare parts, whose cumulative distribution $F_T(t)$ ($t < (T_0, U_0)$) can be gained as

$$F_T(t; r) = \begin{cases} P\{X < t | t < T_0\} = \begin{cases} \frac{F(t; r)}{F(T_0; r)}, & t < T_0 \\ 1, & t \geq T_0 \end{cases} & r \leq r_0 \\ P\left\{X < t \mid t < \frac{U_0}{r}\right\} = \begin{cases} \frac{F(t; r)}{F\left(\frac{U_0}{r}; r\right)}, & t < \frac{U_0}{r} \\ 1, & t \geq \frac{U_0}{r} \end{cases} & r > r_0 \end{cases} \quad (5)$$

Thus, for a given n , it can be got

$$P\{T_S < t|n\} = \begin{cases} F_T^{(S-n)}(t - n \cdot T_0; r) * F(t - n \cdot T_0; r), & r \leq r_0 \\ F_T^{(S-n)}\left(t - n \cdot \frac{U_0}{r}; r\right) * F\left(t - n \cdot \frac{U_0}{r}; r\right), & r > r_0 \end{cases} \quad (6)$$

Based on the convolution theorem and Fourier transform, given inventory level S , age replacement interval (T_0, U_0) , and usage rate r , the cumulative distribution function for T_S can be expressed as

$$G(t; r) = \begin{cases} \sum_{n=0}^S \left\{ \frac{F_T^{(S-n)}(t - n \cdot T_0; r) * F(t - n \cdot T_0; r)}{C_S^n \cdot R^n(T_0; r) \cdot F^{S-n}(T_0; r)} \right\}, & r \leq r_0, \\ \sum_{n=0}^S \left\{ \frac{F_T^{(S-n)}\left(t - n \cdot \frac{U_0}{r}; r\right) * F\left(t - n \cdot \frac{U_0}{r}; r\right)}{C_S^n \cdot R^n\left(\frac{U_0}{r}; r\right) \cdot F^{S-n}\left(\frac{U_0}{r}; r\right)} \right\}, & r > r_0 \end{cases} \quad (7)$$

3.3.2 Cost and probability models in the situation of adequate inventory

As analyzed in Section 3.2, the age replacement cost, failure replacement cost, storage cost, and ordering cost are generated in the situation of adequate inventory. Then the total cost per unit time for a given r is

$$EC_1(T_0, U_0, t_0, S; r) = \left(\frac{C_{m1} + C_{o1} + C_{s1} + C_p}{t_0} \right) \cdot [1 - G(t_0; r)] \quad (8)$$

where C_{m1} , C_{o1} , C_{s1} denote the maintenance cost, ordering cost, and storage cost within ordering period t_0 , respectively.

If the spare parts stock is sufficient, we have C_{m1} as

$$C_{m1} = \begin{cases} \frac{C_p \cdot R(T_0; r) + C_f \cdot F(T_0; r)}{\int_0^{T_0} R(t; r) dt} \cdot t_0, & r \leq r_0 \\ \frac{C_p \cdot R\left(\frac{U_0}{r}; r\right) + C_f \cdot F\left(\frac{U_0}{r}; r\right)}{\int_0^{\frac{U_0}{r}} R(t; r) dt} \cdot t_0, & r > r_0 \end{cases} \quad (9)$$

Let k denotes the expected spare parts consumption within ordering period t_0 ($k \leq S$), and d the ordering cost of a spare parts, added with another spare part at t_0 for system renewal, it's clear that

$$C_{o1} = d \cdot E(k+1) = d + d \cdot E(k) \quad (10)$$

For k can be varied from 0 to S , so

$$E(k) = \sum_{k=0}^S k \cdot P(k; r) \quad (11)$$

$$P(k; r) = \begin{cases} \sum_{n=0}^k C_k^n \cdot R^n(T_0; r) \cdot F^{k-n}(T_0; r), & r \leq r_0 \\ \sum_{n=0}^k C_k^n \cdot R^n\left(\frac{U_0}{r}; r\right) \cdot F^{k-n}\left(\frac{U_0}{r}; r\right), & r > r_0 \end{cases} \quad (12)$$

Substitute equations (13), (14) into (12), C_{o1} can be expressed as

$$C_{o1} = \begin{cases} d + d \cdot \sum_{k=0}^S \left[k \cdot \sum_{n=0}^k C_k^n \cdot R^n(T_0; r) \cdot F^{k-n}(T_0; r) \right], & r \leq r_0 \\ d + d \cdot \sum_{k=0}^S \left[k \cdot \sum_{n=0}^k C_k^n \cdot R^n\left(\frac{U_0}{r}; r\right) \cdot F^{k-n}\left(\frac{U_0}{r}; r\right) \right], & r > r_0 \end{cases} \quad (13)$$

Let n denotes the times of age replacement within ordering period t_0 ($n \leq k$), Δz the actual life for a spare part, h the storage cost of a spare part per unit time, then storage cost within ordering period can be expressed as

$$C_{s1} = h \cdot E \left[\left(\frac{S + S - k}{2} \right) \cdot \Delta z \right] \quad (14)$$

$$\text{where } \Delta z = \begin{cases} \int_0^{T_0} R(T_0; r) dt, & r \leq r_0 \\ \int_0^{U_0/r} R(U_0/r; r) dt, & r > r_0 \end{cases} \quad (15)$$

$$E \left(\frac{S + S - k}{2} \right) = \sum_{k=0}^S \left(\frac{S + S - k}{2} \right) \cdot P(k; r) \quad (16)$$

For n can be varied from 0 to K , so

$$P(k; r) = \begin{cases} \sum_{n=0}^k C_k^n \cdot R^n(T_0; r) \cdot F^{k-n}(T_0; r), & r \leq r_0 \\ \sum_{n=0}^k C_k^n \cdot R^n(U_0/r; r) \cdot F^{k-n}(U_0/r; r), & r > r_0 \end{cases} \quad (17)$$

Thus C_{s1} can be gained as

$$C_{s1} = \begin{cases} h \cdot \sum_{k=0}^S \left[\left(\frac{S + S - k}{2} \right) \cdot \sum_{n=0}^k C_k^n \cdot R^n(T_0; r) \cdot F^{k-n}(T_0; r) \right] \cdot \int_0^{T_0} R(T_0; r) dt, & r \leq r_0 \\ h \cdot \sum_{k=0}^S \left[\left(\frac{S + S - k}{2} \right) \cdot \sum_{n=0}^k C_k^n \cdot R^n(U_0/r; r) \cdot F^{k-n}(U_0/r; r) \right] \cdot \int_0^{U_0/r} R(U_0/r; r) dt, & r > r_0 \end{cases} \quad (18)$$

The total cost per unit time for spare parts system in situation 1 can be expressed as

$$\begin{aligned} EC_1(T_0, U_0, t_0, S) &= \int_{r_l}^{r_u} EC_1(T_0, U_0, t_0, S; r) \cdot f_R(r) dr \\ &= \int_{r_l}^{r_0} EC_{1A}(T_0, U_0, t_0, S; r) \cdot f_R(r) dr + \int_{r_0}^{r_u} EC_{1B}(T_0, U_0, t_0, S; r) \cdot f_R(r) dr \end{aligned} \quad (19)$$

From the equations (9)-(19), we can get $EC_{1A}(T_0, U_0, t_0, S; r) =$

$$\begin{aligned} & \left(\frac{C_p \cdot R(T_0; r) + C_f \cdot F(T_0; r)}{\int_0^{T_0} R(t; r) dt} + \frac{d + d \cdot \sum_{k=0}^S \left[k \cdot \sum_{n=0}^k C_k^n \cdot R^n(T_0; r) \cdot F^{k-n}(T_0; r) \right]}{t_0} \right) \\ & \left(\frac{+ h \cdot \sum_{k=0}^S \left[\left(\frac{S + S - k}{2} \right) \cdot \sum_{n=0}^k C_k^n \cdot R^n(T_0; r) \cdot F^{k-n}(T_0; r) \right] \cdot \int_0^{T_0} R(t; r) dt}{t_0} + \frac{C_p}{t_0} \right) \\ & \cdot \left\{ 1 - \sum_{n=0}^S \left\{ \left[F_T^{(S-n)}(t_0 - n \cdot T_0; r) * F(t_0 - n \cdot T_0; r) \right] \cdot C_S^n \cdot R^n(T_0; r) \cdot F^{S-n}(T_0; r) \right\} \right\} \end{aligned} \quad (20)$$

$$EC_{1B}(T_0, U_0, t_0, S; r) =$$

$$\left(\frac{C_p \cdot R\left(\frac{U_0}{r}; r\right) + C_f \cdot F\left(\frac{U_0}{r}; r\right)}{\int_0^{\frac{U_0}{r}} R(t; r) dt} + \frac{d + d \cdot \sum_{k=0}^S \left[k \cdot \sum_{n=0}^k C_k^n \cdot R^n\left(\frac{U_0}{r}; r\right) \cdot F^{k-n}\left(\frac{U_0}{r}; r\right) \right]}{t_0} \right) \quad (21)$$

$$+ h \cdot \frac{\sum_{k=0}^S \left[\left(\frac{S + S - k}{2} \right) \cdot \sum_{n=0}^k C_k^n \cdot R^n\left(\frac{U_0}{r}; r\right) \cdot F^{k-n}\left(\frac{U_0}{r}; r\right) \right] \cdot \int_0^{\frac{U_0}{r}} R(t; r) dt}{t_0} + \frac{C_p}{t_0}$$

$$\cdot \left\{ 1 - \sum_{n=0}^S \left\{ \left[F_T^{(S-n)}\left(t_0 - n \cdot \frac{U_0}{r}; r\right) * F\left(t_0 - n \cdot \frac{U_0}{r}; r\right) \right] \cdot C_S^n \cdot R^n\left(\frac{U_0}{r}; r\right) \cdot F^{S-n}\left(\frac{U_0}{r}; r\right) \right\} \right\}$$

Substitute equations (20), (21) into (19), $EC_1(T_0, U_0, t_0, S)$ can be got.

3.3.3 Cost and probability models in the situation of insufficient inventory

In the situation of insufficient inventory, taking stock shortage loss into account, the total cost per unit time for a given r is

$$EC_2(T_0, U_0, t_0, S; r) = \left(\frac{C_{m2} + C_{o2} + C_{s2} + C_{los} + C_p}{t_0} \right) \cdot G(t_0; r) \quad (22)$$

where C_{m2} , C_{o2} , C_{s2} , C_{los} denote the maintenance cost, ordering cost, and storage cost within ordering period t_0 , respectively.

Let T_α denotes the expected time of stock shortage within t_0 , then it can be expressed as

$$T_\alpha = E\{t_0 - T_s | T_s < t_0\} = \int_0^{t_0} (t_0 - T_s) dG(t_0 - T_s; r) = \frac{\int_0^{t_0} G(t; r) dt}{G(t_0; r)} \quad (23)$$

For the time of stock shortage within t_0 is T_α , so

$$C_{m2} = (t_0 - T_\alpha) \cdot C_m(T_0, U_0; r)$$

$$= \begin{cases} \frac{C_p \cdot R(T_0; r) + C_f \cdot F(T_0; r)}{\int_0^{T_0} R(t; r) dt} \cdot \left(t_0 - \frac{\int_0^{t_0} G(t; r) dt}{G(t_0; r)} \right), & r \leq r_0 \\ \frac{C_p \cdot R\left(\frac{U_0}{r}; r\right) + C_f \cdot F\left(\frac{U_0}{r}; r\right)}{\int_0^{\frac{U_0}{r}} R(t; r) dt} \cdot \left(t_0 - \frac{\int_0^{t_0} G(t; r) dt}{G(t_0; r)} \right), & r > r_0 \end{cases} \quad (24)$$

$$C_{s2} = h \cdot \frac{S}{2} \cdot (t_0 - T_\alpha) \quad (25)$$

$$C_{o2} = d \cdot (S + 1) \quad (26)$$

$$C_{los} = p \cdot T_\alpha \quad (27)$$

Where p denotes the expected stock shortage loss per unit time.

The total cost per unit time for spare parts system in situation 2 can be expressed as

$$EC_2(T_0, U_0, t_0, S) = \int_{r_l}^{r_u} EC_2(T_0, U_0, t_0, S; r) \cdot f_R(r) dr$$

$$= \int_{r_l}^{r_0} EC_{2A}(T_0, U_0, t_0, S; r) \cdot f_R(r) dr + \int_{r_0}^{r_u} EC_{2B}(T_0, U_0, t_0, S; r) \cdot f_R(r) dr$$

(28)

From the equations (29)-(32), we can get

$$EC_{2A}(T_0, U_0, t_0, S; r) = \left(\left(1 - \frac{\int_0^{t_0} G(y; r) dy}{t_0 \cdot G(t_0; r)} \right) \cdot \frac{C_p \cdot R(T_0; r) + C_f \cdot F(T_0; r)}{\int_0^{T_0} R(t; r) dt} + \frac{d \cdot (S+1)}{t_0} \right) + \frac{h \cdot S}{2} \cdot \left(1 - \frac{\int_0^{t_0} G(y; r) dy}{t_0 \cdot G(t_0; r)} \right) + p \cdot \frac{\int_0^{t_0} G(y; r) dy}{t_0 \cdot G(t_0; r)} + \frac{C_p}{t_0} \cdot \sum_{n=0}^S \left\{ F_T^{(S-n)}(t_0 - n \cdot T_0; r) * F(t_0 - n \cdot T_0; r) \right\} \cdot C_S^n \cdot R^n(T_0; r) \cdot F^{S-n}(T_0; r) \} \quad (29)$$

$$EC_{2B}(T_0, U_0, t_0, S; r) = \left(\left(1 - \frac{\int_0^{t_0} G(y; r) dy}{t_0 \cdot G(t_0; r)} \right) \cdot \frac{C_p \cdot R\left(\frac{U_0}{r}; r\right) + C_f \cdot F\left(\frac{U_0}{r}; r\right)}{\int_0^{\frac{U_0}{r}} R(t; r) dt} + \frac{d \cdot (S+1)}{t_0} \right) + \frac{h \cdot S}{2} \cdot \left(1 - \frac{\int_0^{t_0} G(y; r) dy}{t_0 \cdot G(t_0; r)} \right) + p \cdot \frac{\int_0^{t_0} G(y; r) dy}{t_0 \cdot G(t_0; r)} + \frac{C_p}{t_0} \cdot \sum_{n=0}^S \left\{ F_T^{(S-n)}\left(t_0 - n \cdot \frac{U_0}{r}; r\right) * F\left(t_0 - n \cdot \frac{U_0}{r}; r\right) \right\} \cdot C_S^n \cdot R^n\left(\frac{U_0}{r}; r\right) \cdot F^{S-n}\left(\frac{U_0}{r}; r\right) \} \quad (30)$$

Substitute equations (29), (30) into (28), $EC_2(T_0, U_0, t_0, S)$ can be got.

Finally, integrating the equations above into equation (1), $EC(T_0, U_0, t_0, S)$ is gained as the expected total cost per unit time for spare parts system with two-dimensional maintenance interval (T_0, U_0) and inventory control policy (t_0, S) .

4. Case study

To illustrate the proposed model, we present a case study whose data are originated from Q. W. Hu etc (2015), who have collected two years' operational and maintenance records from a car service shop. The spare parts system for oil filter is studied here to demonstrate the applicability and validity of jointly optimization modelling. For the realization of the car service shop with a cost effective and efficient logistics support, the preventive maintenance and spare parts inventory control must be considered together. The failure time and usage rate of oil filter follow Weibull distribution, and detailed parameters for oil filter spare parts system are listed in Table 1.

Table 1. Case study parameters

parameter	C_p	C_f	h	d	p	Life distribution		Usage rate distribution	
						α_t	β_t	α_r	β_r
value	100	800	50	200	500	3.24	6.85	2.86	4.5×10^3
unit	CNY	CNY	CNY/day	CNY	CNY/day	--	month	--	km/month

According to the iterative algorithm, firstly the interval (T_0, U_0) given fixed t_{0i} and S_j could be optimized, which minimizes the value of $EC(T_0, U_0, t_{0i}, S_j)$. Then, with the iteration of t_0 and S , the

corresponding matrix for $EC(T_0, U_0, t_{0i}, S_j)$ can be got, and the optimal total cost per unit time for oil filter spare parts system $EC(T_0^*, U_0^*, t_0^*, S^*)$ can be gained as 125.4298 CNY/day, with $(T_0^*, U_0^*) = (150 \text{ days}, 4600 \text{ kms})$ and $(S^*, t_0^*) = (6, 300 \text{ days})$, as shown in Figure 4.

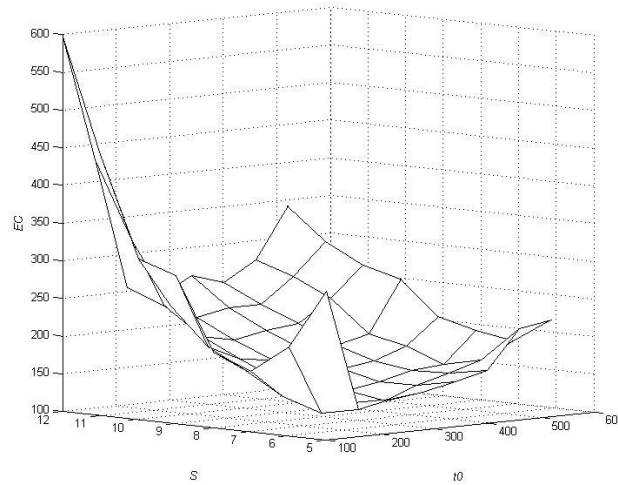


Figure 4. The three-dimensional chart for optimizing (S, t_0) and (T_0, U_0)

Thus, through above analysis and calculation, the total cost per unit time for oil filter spare parts system can be minimized. And it can also be seen that: for different values of (T_0, U_0) and (S, t_0) , the result for $EC(T_0, U_0, t_0, S)$ varies widely, which verifies the necessity and validity of joint optimization modelling.

5. Conclusions and future research

In this paper, for items whose degradation and maintenance are measured both by calendar and usage time, we present a joint optimization model for two-dimensional age replacement interval and spare parts provisioning. A numerical iteration method for model calculation is introduced, and a case study is also given to demonstrate the model's applicability and validity. It can be concluded that, it's necessary to consider the maintenance strategy and spare part inventory jointly to achieve high availability at minimal cost of spare parts system.

There are several topics for future research as follows:

Firstly, to improve the practicality of modelling, more complex reliability structure (e.g. series system, k-out-of-N systems, warm and cold stand-by system, etc.) should be further studied, and their joint optimization of maintenance strategy and spare part inventory should be analyzed and modelled.

Secondly, the approach for modelling two-dimensional maintenance in this paper is based on the assumption that the item usage rate is constant for a given customer. As mentioned in Section 2.2, there still are other approaches which are worthy of further study and comparison.

Lastly, the algorithm for model optimization is given in this paper; however, the computational efficiency isn't so satisfactory. So the related intelligent algorithm, such as genetic algorithm, ant colony optimization, and so on, could be explored in future.

Acknowledgements

Thanks for the support by the National Natural Science Foundation of China (Grant No. 71401173) and the Material Pre-research Foundation (Grant No. 9140A27040314JB34452).

References

- Baik J., Murthy D. N. P., Jack N (2004) Two-dimensional failure modeling with minimal repair. *Naval Research Logistics*, (51):345–362.
- Basten R.J.I., M.C. van der Heijden, J.M.J. Schutten (2012) Joint optimization of level of repair analysis and spare parts stocks. *European Journal of Operational Research*, (222):474-483.

- Chukova S., Johnston M.R. (2006) Two-dimensional warranty repair strategy based on minimal and complete repairs. *Math. Comput. Model*, 44:1133-1143.
- Corbu D., Chukova S., Sullivan J.O. (2008) Product warranty: modelling with 2D-renewal process. *Int. J. Reliab. Saefy*, 2(3):209-220.
- Darko Louit, Rodrigo Pascual, Dragan Banjevic, Andrew K. S. Jardine. (2011) Condition-based spares ordering for critical components. *Mechanical Systems and Signal Processing*, (25):1837–1848.
- Deniz Caglar, Li Chung-Lun, David Simchi-Levi (2004) Two-echelon spare parts inventory system subject to a service constraint. *IIE Transactions Institute of Industrial Engineers*, 36(7):655-666.
- Duchesne T., Lawless J.F. (2000) Alternative time scales and failure time models. *Lifetime Data Anal*, 6:157-179.
- Gertsbakh I.B, Kordonsky K.B. (1998) Parallel time scales and two-dimensional manufacturer and individual customer warranties. *IIE Trans*, 30:1181-1189.
- Hu Qiwei, Bai Yongsheng, Zhao Jianmin. (2015) Modeling spare parts demands forecast under two-dimensional preventive maintenance policy. *Mathematical Problems in Engineering*, article ID 728241, 9 pages. <http://dx.doi.org/10.1155/2015/728241>.
- Jack N., Iskandar B.P, Murthy D.N.P. (2009) A repair-replace strategy based on usage rate for items sold with a two-dimensional warranty. *Reliab. Eng. Syst. Safety*, 94:611-617.
- Jerry Lawless, Joan Hu, Jin Cao. (1995) Methods for the estimation of failure distributions and rates from automobile warranty data. *Lifetime Data Analysis*, (1):227–240.
- Jung M., Bai D. S. (2007) Analysis of field data under two-dimensional warranty. *Reliability Engineering and System Safety*, (92):135–143.
- Karin S. de Smidt-Destombes, Matthieu C. van der Heijden, Aart van Harten. (2009) Joint optimisation of spare part inventory, maintenance frequency and repair capacity for k -out-of- N systems. *International Journal of Production Economics*, 118: 260-268.
- Kennedy W. J., Patterson J.W. etc. (2002) An overview of recent literature on spare parts inventories. *International Journal of Production Economics*, 76(2): 201-215.
- Kim H.G, Rao B.M. (2000) Expected warranty cost of two-attribute free-replacement warranties based on a bivariate exponential distribution. *Comput Ind Eng*, 38:425-434.
- Mahmood Shafiee, Stefranka Chukova etc. (2011) Two-Dimensional Warranty Cost Analysis for Second-Hand Products. *Communications in Statistics — Theory and Methods*, 40: 684-701.
- Manna D.K, Pal S., Sinha S. (2006) Optimal determination of warranty region for 2D policy: a customers' perspective. *Comput. Indust. Eng*, 50:161-174.
- Martorell S., S´anchez A., Carlos S., Serradell V. (2002) Comparing effectiveness and efficiency in technical specifications and maintenance optimization. *Reliability Engineering and System Safety*, 77(3):281–289.
- Pal S., Murthy G.S.R. (2003) An application of Gumbel's bivariate exponential distribution in estimation of warranty cost of motor cycles. *Int J Qual Reliab Manage*, 20: 488-502.
- Rai B., Singh N. (2005) A modeling framework for assessing the impact of new time/mileage warranty limits on the number and cost of automotive warranty claims. *Reliab. Eng. Syst. Safety*, 88:157-169.
- Seongmin Moon, Christian Hicks, Andrew Simpson. (2012) The development of a hierarchical forecasting method for predicting spare parts demand in the South Korean Navy—A case study. *Int. J. Production Economics*, (140):794-802.
- Tom Chen, Elmira Popova. (2002) Maintenance policies with two-dimensional warranty. *Reliability Engineering and System Safety*, (77):61–69.
- Tsai Y.T., Wang K.S., Tsai L.C. (2004) A study of availability-centered preventive maintenance for multi-component systems. *Reliability Engineering and System Safety*, 84(3): 261–270.
- Wang Wenbin. (2011) A joint spare part and maintenance inspection optimisation model using the Delay-Time concept. *Reliability Engineering and System Safety*, 96: 1535-1541.
- Xu Chang an etc. (2011) Optimal joint spare part stocking and age based replacement policy. *Proceedings of ICQR2MSE*, vol. 2: 636-639.
- Yang S.C, Nachlas J.A. (2001,) Bivariate reliability and availability modelling. *IEEE Trans Reliab*, 50:26-35.

Calendar-based age replacement policy and its generalizations

Dragan Banjevic* and Maliheh Aramon Bajestani**

Department of Mechanical & Industrial Engineering, University of Toronto, Toronto, Ontario, Canada

**banjev@mie.utoronto.ca, ** maramon@mie.utoronto.ca*

Abstract. We introduce an age-based replacement policy in which the preventive replacements are restricted to specific calendar times. Under this policy, the assets are renewed at failure or if their ages are greater than or equal to a replacement age at given calendar times, whichever occurs first. This policy is logistically applicable in industries such as utilities where there are large and geographically diverse populations of deteriorating assets with different installation times. This policy incorporates the classical age and block policies as a special case. The classical renewal reward theorem cannot be applied directly to find the cost per time unit since the renewal cycles are dependent random variables. Using the theory of Markov chains with general state space and a suitably defined ergodic measure, we find the optimal replacement age that minimizes the long-run expected cost per time unit and show that it can be found using the expected cost and length of one renewal cycle where the expectation is calculated with respect to forward time ergodic measure. We further show how the ergodic measure can be used to analyze more complicated policies, such as ones introduced by Scarf and Dearn in 2003. We numerically illustrate some of our results using a real data set of utility wood poles' maintenance records.

1. Introduction

The most classical maintenance policies in the reliability and maintenance literature are the age-based replacement policy and the block replacement policy (Barlow and Proschan, 1996). In the age policy, a component is replaced with a new one at failure or at age t_p , whichever occurs first. In the block policy (a “calendar-based” policy), the component is replaced at failure and at equally spaced time points independent of the failure history of the component. Both policies are extensively studied in the literature, considering more general models and cost structures. The literature on maintenance policies is very extensive. A comprehensive review of the literature on maintenance policies can be found in the works by Wang (2002), and Nakagawa (2008).

In the age-based replacement policy and its extensions, the next replacement age is measured from the time of the last replacement or the last repair. The preventive replacements cannot be planned several steps in advance and the maintenance logistics should be available at any calendar time to replace the component reaching its replacement age. Therefore, implementing the age-based replacement policy and its generalizations is not logistically convenient in asset-intensive industries such as utilities. In this paper, we introduce an alternative policy, called “calendar-based age policy”. In this policy, the assets are replaced with new ones at failure or if their ages are greater than or equal to the predetermined replacement age at given checking points, whichever occurs first. The policy is motivated by a population of several thousand utility wood poles in the distribution system of a Canadian electricity distributor. The company already has treatment schedules for poles installed in different locations. Therefore, using the calendar-based age replacement policy where poles can be preventively replaced at already fixed scheduled treatment points is practically justifiable on the grounds of administrative and logistical feasibility for the company.

The primary difference of the calendar-based age replacement policy from the age-based replacement policy is that the checking points schedule is based on fixed calendar times, not on the time since previous replacement (renewal). Therefore, the renewal cycles (the times between consecutive replacements) are dependent random variables. The dependency of the renewal cycles under a fixed schedule was discussed by Wang et al. (2010) and Wang and Banjevic (2012). Wang et al. (2010) study a multi-component delay time model where the inspections are scheduled according to a fixed plan for the whole system. They define the forward time as the time to failure since the previous inspection. Assuming that the limiting distribution of the forward time exists, they derive the unconditional expected cost and unconditional expected length of the renewal cycle. Following the classical renewal reward theorem, Wang et al. (2010) use the unconditional expected cost per one renewal cycle divided by the unconditional expected length of one cycle as the

objective function. However, since the renewal cycles are dependent under a fixed inspection program, the renewal reward theorem cannot be formally applied.

Our calendar-based age replacement policy is also a generalization of the modified block replacement policy introduced by Berg and Epstein (1976), in which the components that were recently replaced due to failure are not replaced at preventive replacement points. In the calendar-based age replacement policy the threshold replacement age can be any positive value, shorter or longer than the length of the scheduled checking interval. Berg and Epstein (1976), as well as the subsequent papers (Archibald and Dekker, 1996; Scarf and Dears, 2003; Li, 2005) heuristically derive the expected cost per time unit for an infinite time horizon as if the replacement points were independent renewal points. In this paper, we present the “formal” analysis of the calendar-based age replacement policy through the theory of Markov chains with general state space and distinguish between dependent renewal cycles and independent regeneration cycles. Using the ergodicity of a Markov chain with a general state space, we show how the expected cost and length of a regeneration cycle can be calculated using the expected cost and length of one renewal cycle. The Markov renewal stochastic technique with general state space has been used by some authors in the maintenance literature for the analytical analysis of the maintenance policies which could only be analyzed through simulation (see, e.g., Mercier and Pham, 2014). Similar to these works, our paper shows that the Markov renewal stochastic technique is more powerful than the classical renewal technique. It also allows us to easily extend our results, analyzing more complicated policies (e.g., of Scarf and Dears, 2003), where the solutions of heuristic analysis would be difficult to justify. This technique also allows for calculating the other useful operating characteristics of the calendar-based age replacement policy such as the limiting distributions of the backward and the forward times (not presented in this paper).

2. Problem formulation using a Markov chain with a general state space

We consider one component system, failing according to the probability density function $f(x)$, with reliability function $R(x)$. To avoid technical difficulties, we assume $f(x) > 0, x > 0$. In the calendar-based age replacement policy, the age of the component is checked at equidistant calendar times $\Delta, 2\Delta, \dots$, called “checking points” and the component is renewed if its age is greater than or equal to t_p or at failure, whichever occurs first. There are two types of renewals: renewal at failure and renewal at checking point. The former is a failure renewal and the latter is a preventive renewal. The goal is to find the optimal replacement age t_p that minimizes the expected cost per time unit on long run. The components of the cost are c_s , c_f , and c_p denoting the cost of checking the age (inspection), the failure renewal cost, and the preventive renewal cost, respectively. c_s can be negligible if the age is measured in calendar time scale and there is a computerized system recording the installation time and the failure history of the components. On the other hand, it can be high if the age is measured on a deterioration scale and a site visit and specific measurements and equipments are required to retrieve the information on the deterioration levels of the components. In the context of our example, the poles are visited for regular treatment regardless of their ages every 10 years, which contributes to c_s cost.

Due to the fixed schedule of preventive replacement intervals, the probability that the next renewal will be at failure or at a checking point depends on whether the previous renewal was after failure or a preventive one. Therefore, the sequence of cycle lengths is not a renewal process. To properly formulate the calendar-based age replacement policy, we will construct a renewal process by defining a suitable Markov chain.

Let $\{X_n : n = 0, 1, 2, \dots\}$ be a homogeneous Markov chain on a measurable space $(S, \mathcal{B})$ with transition function $P(x, E : t_p, \Delta) = P\{X_{n+1} \in E \mid X_n = x, t_p, \Delta\}$, $x \in S, E \in \mathcal{B}$ where $S = [0, \Delta)$ and $\mathcal{B}$ is a Borel σ -field on S . In the calendar-based age replacement policy, we define X_n as the n -th forward time being equal to the time between the n -th renewal and its preceding checking point. Alternatively, we could define X_n to represent the age at the beginning of the n -th checking point, as it has been used heuristically by Berg and Epstein (1976) and other authors, but it appeared to be computationally more complicated. If the n -th renewal occurs at one of the checking points, $X_n = 0$; otherwise, $0 < X_n < \Delta$. Without loss of generality, we assume that the fixed calendar-based checking points schedule starts before the installation of the component. Since the age of the component at some checking point might be greater than or equal to t_p , there is a positive probability that the chain's state will be equal to 0, representing the preventive replacement. Let

$$p(x, 0; t_p, \Delta) = P\{X_{n+1} = 0 \mid X_n = x, t_p, \Delta\}, P(x, E; t_p, \Delta) = \int_E p(x, y; t_p, \Delta) dy, E \subset S, E \in \mathcal{B}.$$

The one-step transition kernel of the Markov chain is

$$p(x, 0) = p(x, 0; t_p, \Delta) = R(n(x)\Delta - x), y = 0, 0 \leq x < \Delta,$$

$$p(x, y) = p(x, y; t_p, \Delta) = f(x - y)I(x \leq y) + \sum_{n=1}^{n(x)-1} f(n\Delta + y - x), 0 < y, 0 \leq x < \Delta,$$

where $I(A)$ denotes the indicator function of an event A , and $n(x)$ denotes the smallest integer greater than or equal to $(t_p + x)/\Delta$. Obviously, $P(x, S) = p(x, 0) + \int_0^\Delta p(x, y) dy = 1, x \in S$.

We need also two random sequences, L_n and C_n , the random length and the random cost associated with the $n + 1$ -st renewal cycle, respectively, $n = 0, 1, 2, \dots$. The two sequences are independent conditionally on a realization of $X_n = x_n, n = 1, 2, \dots$. Their distributions depend only on x_n . That is, for L_n (or C_n), $P\{L_n \in A \mid x_0, x_1, \dots, x_n\} = P\{L_n \in A \mid x_n\}, A \subset [0, \infty)$. Let also $V_i = \sum_{n=0}^{i-1} L_n, i \geq 1, V_0 = 0$. If the cost of each renewal cycle is paid at the end of the cycle, the total cost up to time t is $C(t) = \sum_{n=0}^{i-1} C_n, V_i \leq t < V_{i+1}$. The optimal t_p^* is t_p that minimizes the cost per time unit on the long run, $\lim_{t \rightarrow \infty} (E[C(t)]/t)$. However, in some applications Δ can also be considered as a decision variable and the optimization problem is in finding optimal t_p and Δ , assuming the basic costs, such as c_s , depend on Δ .

It is intuitively clear whenever a renewal occurs at a checking point, the Markov chain restarts itself, or the preventive renewal is the regeneration point. Let $N_1 = \inf\{n: X_n = 0, n \geq 1\}$ be a random variable denoting the number of renewals until the first preventive renewal, and $T_1 = \sum_{n=0}^{N_1-1} L_n$ be the total time until the first preventive renewal. Let $T_k = \sum_{n=N_{k-1}}^{N_k-1} L_n, k \geq 2$, be the k -th regeneration time where N_k is the number of renewals until the k -th preventive renewal. The sequence of regeneration times, $T_k, k = 1, 2, \dots$, is a delayed renewal process as T_1 has a different distribution than $T_k, k = 2, 3, \dots$. Therefore, using the renewal reward theorem (Rausand and Hoyland, 2004) for delayed process, the expected cost per time unit on long run is

$$\lim_{t \rightarrow \infty} \frac{E[C(t)]}{t} = \frac{E[\sum_{n=N_1}^{N_2-1} C_n]}{E[\sum_{n=N_1}^{N_2-1} L_n]} = \frac{E[\sum_{n=0}^{N_1-1} C_n \mid X_0 = 0]}{E[\sum_{n=0}^{N_1-1} L_n \mid X_0 = 0]}. \quad (1)$$

To use the above result, we need to prove that the first regeneration time exists and will occur infinitely often. That is, we need to prove that the Markov chain $X_n, n = 1, 2, \dots$ is ergodic. Then, we need to find a way to calculate the ratio in eq. (1), that is, to calculate $E[\sum_{n=0}^{N_1-1} C_n \mid X_0 = 0]$ and $E[\sum_{n=0}^{N_1-1} L_n \mid X_0 = 0]$ more effectively.

Let us define the n -th step transition kernel by

$$p^{(n)}(x, 0) = P\{X_n = 0 \mid X_0 = x\}, p^{(n)}(x, E) = P\{X_n \in E \mid X_0 = x\} = \int_E p^{(n)}(x, y) dy, E \subset S \setminus \{0\}, E \in \mathcal{B}.$$

If the Markov chain is ergodic, i.e., $\lim_{n \rightarrow \infty} p^{(n)}(x, y) = p(y)$ exists, $p(y)$ will be independent of the initial state x , and will be a proper distribution ($p(0) + \int_0^\Delta p(y) dy = 1$). This gives the following system of equations:

$$p(0) = p(0)p(0, 0) + \int_0^\Delta p(u)p(u, 0) du, p(y) = p(0)p(0, y) + \int_0^\Delta p(u)p(u, y) du, y > 0. \quad (2)$$

To prove the ergodicity of the Markov chain $\{X_n\}$ with a general state space, it suffices to show the chain is recurrent and has an invariant probability measure (Athreya et al., 1978; Athreya and Ney, 1978). According to Athreya et al. (1978), the chain $\{X_n\}$ is $(A, \lambda, \varphi, n_0)$ -recurrent if a set $A \in \mathcal{B}$, a probability measure φ on A , a constant $\lambda > 0$, and an integer n_0 exist, such that:

- (i) $P\{X_n \in A \text{ for some } n \geq 1 \mid X_0 = x\} = 1, \forall x \in S$,
- (ii) $P\{X_{n_0} \in B \mid X_0 = x\} \geq \lambda \varphi(B), \forall x \in A, B \subset A, B \in \mathcal{B}$.

An obvious candidate for the regeneration set A in our problem is a single point $x = 0$, since the component's age eventually will be greater than or equal to the replacement age t_p at some checking point and a preventive renewal will then occur. Considering $A = \{0\}, \varphi(0) = 1, n_0 = 1$, and $0 < \lambda \leq R(t_p + 2\Delta)$, it can be proved that our chain $\{X_n\}$ is recurrent. Based on Theorem 2.1 of Athreya et al. (1978), $\nu(E) = E[\sum_{n=0}^{N_1-1} I(X_n \in E) \mid X_0 = 0]$ is an invariant and unique up to a multiplicative constant measure for the recurrent Markov chain, where N_1 is a random integer denoting the number of renewals until the chain enters the regeneration set $\{0\}$ for the first time. Furthermore, the Corollary 2.2 of Athreya et al. (1978) states an invariant probability measure $\pi(E) = \nu(E)/\nu(S)$ exists if and only if $\nu(S) = E[N_1 \mid X_0 = 0] < \infty$, which can be easily verified in our case, under the above assumptions. It can be shown that $\pi(E)$ in our problem is

absolutely continuous in $(0, \Delta)$ and has a probability density function satisfying the integral equations (2). Furthermore, it can be shown that $\nu(0)=1$ and $\nu(E)=\int_E g(x)dx, E \subset S \setminus \{0\}$, where $g(x)=p(x)/p(0)$, and $g(0)=1$. Equations (2) can be simplified to include $g(x)$ only, just by dividing by $p(0)$. To solve modified equations (2), we can use the successive approximation procedure, where we start for $y > 0$ with $g_1(y)=p(0, y)$, $g_i(y)=p(0, y) + \int_0^\Delta g_{i-1}(u)p(u, y)du$, or by using an equivalent system of linear equations from discretized state space and applying the quadrature rules. After solving the modified equations (2), we can calculate $p(0)=1/[1+\int_0^\Delta g(u)du]$ and $p(y)=g(y)p(0)$, $0 < y < \Delta$. We use $\varepsilon=0.001$ in our numerical study in Section 4. Once we find the ergodic measure, we can calculate the expected cost per time unit as

$$\lim_{t \rightarrow \infty} \frac{E[C(t)]}{t} = \frac{E[C_0 | X_0 = 0] + \int_{S \setminus \{0\}} E[C_0 | X_0 = y]g(y)dy}{E[L_0 | X_0 = 0] + \int_{S \setminus \{0\}} E[L_0 | X_0 = y]g(y)dy}, \quad (3)$$

where the numerator and the denominator are the expected cost and the expected length of one regeneration cycle, respectively.

By considering the time to first failure, we can show that,

$$E[C_0 | X_0 = y] = c_f(1 - R(n(y)\Delta - y)) + c_p R(n(y)\Delta - y) + c_s \sum_{n=1}^{n(y)} R(n\Delta - y), \quad (4)$$

$$E[L_0 | X_0 = y] = \int_0^{n(y)\Delta - y} R(z)dz, y > 0. \quad (5)$$

It should be noted that the ergodic measure $p(\cdot)$, the expected cost and mean time of one renewal cycle are functions of t_p . Therefore, the optimal replacement age t_p is hard to find directly, except by numerical search; the optimal solution to the “classical” age policy may be a good starting point. An alternative, computationally simpler formula for the cost per time unit of the modified block replacement policy of Berg and Epstein (1976) can be obtained from (3) - (5), for $0 < t_p < \Delta$, after some calculation, as

$$\frac{c_p + c_f \int_0^\Delta g(y)dy + c_s \int_{\Delta - t_p}^\Delta R(\Delta - y)g(y)dy}{t[1 + \int_{\Delta - t_p}^\Delta R(\Delta - y)g(y)dy]}. \quad (6)$$

The limiting distributions of the backward and forward times are also of interest. Assuming t as a current time, the time between the latest renewal before t and t represents the current age of the component (the backward recurrence time), and the time between t and the next renewal immediately after t represents the remaining life of the component (the forward recurrence time). For example, in our case, the backward time distribution provides information on the current ages of the components in the fleet and the forward time distribution can be used to find the expected number of renewals in future intervals allowing to plan the budget. It can be shown that these two limiting distributions have different asymptotic behaviors; they also depend on the time since previous checking point. Details are not given here.

3 Generalizations of the calendar-based age policy

Prolonged Downtime. In the calendar-based age replacement policy, the component is immediately replaced with a new one at failure. However, if the failure happens at most t_d time units before one of the fixed checking points, $0 < t_d < \Delta$, it might be more reasonable to delay the failure renewal until the scheduled checking point, by paying some downtime cost. We may define X_n as the forward time of the new policy, where $0 \leq X_n < \Delta - t_d$. The renewal at checking points, where $X_n = 0$, is the regeneration point.

Alternating Fixed Checking Intervals. In some cases it may be more realistic to have unequal checking intervals, such as depending on season and location. We may consider a policy where the length of checking intervals alternates between short and long values, Δ_s and Δ_l . The component is replaced with a new one at failure or if its age is greater than or equal to t_p at a checking point, whichever occurs first. We can define a two-dimensional homogeneous Markov chain $\{(X_n, Y_n)\}$ with a state space $S = [0, \max(\Delta_s, \Delta_l)] \times \{0, 1\}$, where X_n is the n -th forward time, and Y_n the type of the previous checking interval at the n -th renewal, being equal to 0 if it is short and 1, if it is long. If we define a preventive renewal at the short interval as the regeneration point, it can be proved that the chain $\{(X_n, Y_n)\}$ is ergodic. The system of integral equations for

the ergodic measure of $\{(X_n, Y_n)\}$ is more involved. The above result can be easily extended to a case with $k > 2$ different alternating checking point intervals.

Random Checking Intervals. We can further generalize to the case that the checking intervals are short or long, at random, and construct a two-dimensional Markov chain as discussed above. The ergodic measure of this policy is slightly different from the policy with alternating fixed checking intervals. Another generalization could be with several checking intervals, or even with their lengths coming from a given distribution, or from a point process (such as a homogeneous/nonhomogeneous Poisson process).

Grouped Modified Block Replacement. Scarf and Deara (2003, policy IVa2) consider a two-component system with type I failure dependence where the failure of component 1 (2) induces a failure of component 2 (1) with probability p (q) and has no effect on component 2 (1) with probability $1 - p$ ($1 - q$). The system fails if at least one of the components fails. Under this policy, the failed component(s) are replaced on the failure of the system. At fixed checking points, scheduled in Δ time units, we also preventively replace components 1 and 2 if their respective ages are greater than their critical replacement ages t_p^1 and t_p^2 . To analyze this policy, we can introduce $\{(X_n, Y_n, A_n)\}$ a three-dimensional homogeneous Markov chain with $0 \leq X_n < \Delta$, $Y_n \in \{0, 1, 2\}$, and $0 \leq A_n \leq \Delta + \max(t_p^1, t_p^2)$, where X_n is the n -th forward time, equal to the time between the previous checking point and the n -th renewal, Y_n is the indicator of replacement type (indicating that both components are replaced, 1st, or 2nd), and A_n is the age of the component that is not replaced at the n -th renewal. It can be proved that the preventive renewal of both components, i.e., the state $(0, 0, 0)$, is the regeneration point. We think that the same three-dimensional Markov chain can be used to analyze the other policies including opportunistic grouped modified block replacement policy (policy IVc2) of Scarf and Deara (2003) which they analyzed using simulation. The Markov renewal technique can also be applied to analyze other block and age-based replacement policies developed by Scarf and Deara (2003) for a two component system with failure dependency. This technique will likely result in simpler expressions for long-run expected cost per time unit whose calculations are easier and faster. We have to admit, that due to increase in dimensionality, some parts of the model, such as the ergodic measure itself, still can be easier to calculate using simulation, but it is for more detailed analysis.

5 Numerical study

We use the utility wood poles example in the distribution system of a Canadian electricity distributor with $\Delta = 10$ years. We assume the time to failure follows a Weibull distribution, with the parameters shape and scale $\hat{\beta} = 5.1$ and $\hat{\eta} = 114.33$ years, estimated from real data.

Optimal Replacement Age. Table 1 (the third and fourth columns) shows the optimal replacement age, t_p^* , in years, and the optimal long-run expected cost per year, γ^* , for the calendar-based age replacement policy for a selection of cost parameters. The preventive renewal cost c_p is set at \$7000, approximately a cost of buying a new pole and its installation. The results are presented here for different ratios of the failure renewal cost to the preventive renewal cost, c_f/c_p , and three different costs of checking the age, c_s . The investigation of the equation for γ^* indicates there are certain intervals for c_s in which the optimal value of the replacement age does not change, if calculated in multiples of one year.

Comparison of the calendar-based age policy and two other age-based policies. We refer to the “classical” age-based replacement policy in which the optimal t_p is a multiple of a fixed Δ , as a *restricted* age-based replacement policy. As we already mentioned, the implementation of the calendar-based age replacement policy is logistically easier than the age-based policies. We compare the calendar-based age replacement policy with the two other policies, assuming they are all feasible. We assume the cost of replacing a pole at failure in both age-based policies (classical and restricted) c_f^a is equal to $c_f \cdot t_p^*$ and γ^* for the two other policies are shown in Table 1 (the sixth to ninth columns) for different ratios of c_f/c_p and c_p^a/c_p . As expected, the classical age-based policy always results in lower long-run expected cost per year than the restricted age-based policy. Also, the restricted and classical age-based replacement policies result in a longer optimal replacement age than the calendar-based age replacement policy for all ratios c_f/c_p , c_p^a/c_p , and c_s values. The results further indicate that the calendar-based age replacement policy is not always optimal when $c_p^a/c_p = 1$ for all c_f/c_p since it results in a slightly bigger long-run expected cost per year. For example, if $c_p^a/c_p = 1$ and $c_f/c_p = 2$, $\gamma^* = 102.2260$ for the restricted age-based policy and $\gamma^* =$

102.0336 for the classical age-based policy, while $\gamma^* = 102.2287$ for the calendar-based age policy. However, if the preventive renewal cost is lower than the cost of preventive replacement, policy. i.e., $c_p^a/c_p > 1$, the long-run expected cost per year of the calendar-based age policy can be significantly lower than the long-run cost of both age-based policies for all three c_s values, making it the optimal policy. For example, when $c_p^a/c_p = 2$ and $c_f/c_p = 4$, $\gamma^* \approx 240$ for both age-based policies while for the calendar-based age policy $\gamma^* \approx 125, 135$, and 175 for $c_s = 0, 100$, and 500 , respectively. It is clear that for $c_p^a/c_p < 1$, the age-based policies would result in a smaller long-run expected cost per year than the calendar-based age replacement policies. However, because of the logistical superiority of the calendar-based age policy, it is more reasonable to assume that $c_p^a/c_p > 1$ in the real-world.

$\frac{c_f}{c_p}$	c_s	Calendar-based Age Policy		$\frac{c_p^a}{c_p}$	Restricted Age-based Policy		Classical Age-based Policy	
		t_p^*	γ^*		t_p^*	γ^*	t_p^*	γ^*
2	0	82	102.2287	1	90	102.2260	87	102.0336
	100	82	112.2226					
	500	82	152.1981					
4	0	65	125.2691	1	70	125.2374	70	125.2374
	100	65	135.2687	2	90	204.4521	87	204.0672
	500	65	175.2670	3	110	254.3774	109	254.3446
6	0	58	138.8642	1	60	138.8739	63	138.1426
	100	58	148.8641	2	80	233.2216	76	231.9474
	500	58	188.8637	3	90	306.6781	87	306.1008
				4	100	361.7791	100	361.7791
				5	120	394.4464	121	394.4272

Table 1: The optimal replacement age, t_p^* years, and the optimal long-run expected cost per year, γ^* , of the calendar-based age replacement policy, the restricted age-based replacement policy, and the classical age-based policy for different ratios of cost parameters when $c_p = \$7000$.

References

- Archibald TW, Dekker R (1996) Modified block replacement for multiple-component systems. *IEEE Transactions on Reliability* **45**, 75–83.
- Athreya KB, McDonald D, Ney P (1978) Limit theorems for semi-Markov processes and renewal theory for Markov chains. *The Annals of Probability* **6**, 788–797.
- Athreya KB, Ney P (1978) A new approach to the limit theory of recurrent Markov chains. *Transactions of the American Mathematical Society* **245**, 493–501.
- Barlow RE, Proschan F (1996) *Mathematical Theory of Reliability*. SAIM.
- Berg M, Epstein B. (1976) A modified block replacement policy. *Naval Research Logistics* **23**, 15–24.
- Li H (2005) Stochastic comparison of age-dependent block replacement policies. *Methodology and Computing in Applied Probability* **7**, 473–488.
- Mercier S, Pham HH (2014) A condition-based imperfect replacement policy for a periodically inspected system with two dependent wear indicators. *Applied Stochastic Models in Business and Industry* **30**, 766–782.
- Nakagawa T (2008) *Advanced Reliability Models and Maintenance Policies*. Springer.
- Rausand M, Hoyland A (2004) *System Reliability Theory*, Wiley.
- Scarf, PA, Dears M (2003) Block replacement policies for a two-component system with failure dependence. *Naval Research Logistics* **50**, 70–87.
- Wang H (2002) A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research* **139**, 469–489.
- Wang W, Banjevic D (2012) Ergodicity of forward times of the renewal process in a block-based inspection model using the delay time concept. *Reliability Engineering and System Safety* **100**, 1–7.
- Wang W, Banjevic D, Pecht M (2010) A multi-component and multi-failure mode inspection model based on the delay time concept. *Reliability Engineering and System Safety* **95**, 912–920.

Parametric bootstrap goodness-of-fit tests for imperfect maintenance models

Cécile Chauvel*, Jean-Yves Dauxois**, Laurent Doyen* and Olivier Gaudoin*

*Laboratoire Jean Kuntzmann, Université Grenoble Alpes, Grenoble, France.

cecile.chauvel@imag.fr, laurent.doyen@iut2.upmf-grenoble.fr, olivier.gaudoin@imag.fr.

**Institut de Mathématiques de Toulouse, Université de Toulouse-INSA, Toulouse, France.

jean-yves.dauxois@insa-toulouse.fr.

Abstract The simultaneous modelling of ageing and maintenance efficiency of repairable systems is a major issue in reliability. Many imperfect maintenance models have been proposed. To analyze a dataset, it is necessary to check whether these models are adapted or not. In this paper, we propose a general methodology for testing the goodness of fit of any kind of imperfect maintenance model. Two families of tests are presented, based respectively on martingale residuals and probability integral transforms. The quantiles of the test statistics distributions under the null hypothesis are computed with parametric bootstrap methods. An extensive simulation study is provided, from which we recommend the use of two tests in practice, one from each family. Finally, the tests are applied to several real datasets.

1. Imperfect maintenance models

Let us assume that we observe the n first failure times of a repairable system $T_1, T_2, \dots, T_n$, with $T_0 = 0$. A corrective maintenance, or repair, is made after each failure. The repair duration is not taken into account, or considered to be negligible. Let $N = (N_t)_{t \geq 0}$ denote the counting process of failures. A stochastic model for this process of recurrent events is completely characterized by the intensity $\lambda = (\lambda_t)_{t \geq 0}$ of the failure counting process defined as

$$\lambda_t = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(N_{t+\Delta t} - N_t = 1 \mid \mathcal{H}_t^-), \quad t \geq 0,$$

where $\mathcal{H}_t^-$ is the past of the process just before t .

A stochastic model for the failure process is composed of two parts. The first part of the model is the initial intensity, which expresses the intrinsic wear of the system before the first maintenance. Usually, the first failure time is assumed to follow a Weibull distribution, so that the initial intensity is the intensity of a Power Law Process (PLP), $h(t) = abt^{b-1}$ ($a > 0, b > 0$), for $t \geq 0$. The intensity of a Log Linear Process (LLP) is also considered in practice: $h(t) = \exp(a + bt)$ ($a, b \in \mathbb{R}$), for $t \geq 0$. In this paper, we assume that the system is wearing, so the initial intensity is an increasing function of time. Then, we will set $b > 1$ for the PLP and $b > 0$ for the LLP.

The second part of the model characterizes the maintenance effect on the system. The first basic assumption is to consider that after maintenance, the system is in the same state as before. This is known as minimal maintenance or As Bad As Old (ABAO) effect. The corresponding random processes are Non-Homogeneous Poisson Processes and the intensity is

$$\lambda_t = h(t). \quad (1)$$

The second basic assumption is to consider that maintenance leaves the system as if it were new. This is known as perfect maintenance or As Good As New (AGAN) effect. The corresponding random processes are Renewal Processes and the intensity is

$$\lambda_t = h(t - T_{N_t^-}). \quad (2)$$

Reality is between these two extreme cases. This situation is known as imperfect maintenance. Many imperfect maintenance models have been proposed. In this paper, we will focus on the following models, defined by their intensity:

- Brown-Proschan (BP) model (Brown and Prochan, 1983) : each maintenance is AGAN with a probability $p \in [0, 1]$ and ABAO with a probability $1 - p$.

$$\lambda_t = h \left(t - T_{N_t^-} + \sum_{j=1}^{N_t^-} \left(\prod_{k=j}^{N_t^-} (1 - B_k) \right) (T_j - T_{j-1}) \right). \quad (3)$$

where the $(B_k)_{k \in \mathbb{N}}$ are i.i.d. random variables with Bernoulli distribution of parameter p , such that $B_k = 1$ if the k th maintenance is AGAN and $B_k = 0$ if it is ABAO.

- Quasi-renewal (QR) model (Wang and Pham, 1996), or Geometric process (Lam and Zhang, 2003) : the times between two successive failures decrease or increase geometrically, depending on whether the system wears or rejuvenates.

$$\lambda_t = q^{-N_t^-} h \left(q^{-N_t^-} (t - T_{N_t^-}) \right). \quad (4)$$

where $q > 0$ is a parameter characterizing the effect of repair.

- The Extended Geometric Process (EGP, Bordes and Mercier, 2013) reduces the geometrical growth of the QR model:

$$\lambda_t = q^{-b_{N_t^-}} h \left(q^{-b_{N_t^-}} (t - T_{N_t^-}) \right). \quad (5)$$

where $(b_i)_{i \in \mathbb{N}^*}$ is a non-decreasing sequence of non-negative real numbers such that $b_1 = 0$ and b_i tends to infinity as i goes to infinity.

The last class of imperfect maintenance models presented in this article is the class of virtual age models (Kijima, 1989). The models of this class depend on a sequence of positive random variables $(A_i)_{i \in \mathbb{N}^*}$ called effective ages. They assume that after the i th failure, the system behaves like a new system which has not failed until A_i , for $i \in \mathbb{N}^*$. The failure intensity is

$$\lambda_t = h \left(A_{N_t^-} + t - T_{N_t^-} \right). \quad (6)$$

A virtual age model is defined by a particular expression of the effective ages. Among them, we will consider the following models.

- The Arithmetic Reduction of Age model with infinite memory (ARA_∞, Doyen and Gaudoin, 2004), or Kijima type II model with deterministic constant effect : the effect of repair is to reduce the effective age of the system by a factor ρ .

$$\lambda_t = h \left(t - \rho \sum_{j=0}^{N_t^- - 1} (1 - \rho)^j T_{N_t^- - j} \right). \quad (7)$$

- The Arithmetic Reduction of Age model with memory 1 (ARA₁, Doyen and Gaudoin, 2004), or Kijima type I model with deterministic constant effect : the effect of repair is to reduce the supplement of age since the last failure by a factor ρ .

$$\lambda_t = h \left(t - \rho T_{N_t^-} \right). \quad (8)$$

In both ARA models, the value of $\rho \in [0, 1]$ represents the effect of repair. If $\rho \in]0, 1[$ the repair is efficient, if $\rho = 1$ the repair is optimal (AGAN), if $\rho = 0$ the repair is minimal (ABAO).

To assess both the intrinsic ageing and the repair effect, one has to estimate the parameters of these imperfect maintenance models. This is usually done by the maximum likelihood method. The likelihood function associated to the observation of the n first failure times $T_1, \dots, T_n$ is

$$L_n = \left(\prod_{i=1}^n \lambda_{T_i} \right) \exp \left(- \int_0^{T_n} \lambda_t dt \right). \quad (9)$$

2. Goodness-of-fit tests for imperfect maintenance models

For each particular application and observed data set, practitioners are confronted to the problem of choosing the “most appropriate” model among the big range of available models. More precisely, practitioners need to evaluate whether the observations could be realizations of random variables following the assumptions of a

given model, or not. This is a classical statistical issue on model selection. A solution is to build a statistical test of hypotheses, called a goodness-of-fit test. Goodness-of-fit tests are tools for checking whether the model fits well the data or not. These tests are well-known for simple situations such as samples: classical techniques are available for determining if a sample can be considered as issued from independent and identically distributed (i.i.d.) random variables. These techniques can be adapted to build goodness-of-fit tests for AGAN models. Testing the ABAO assumption is equivalent to build goodness-of-fit tests for Non-Homogeneous Poisson Processes. To that aim, several methods have been proposed yet. To our knowledge, only Lindqvist et al (2003) and Liu et al (2012) have proposed methods for testing the goodness of fit of imperfect maintenance models. They both consider a transformation of the failure times into uniform random variables. However, these authors did not provide a detailed analysis of their tests, nor performed simulation studies to evaluate their performances.

Given that a point process model is characterized by its intensity, an imperfect maintenance model can be denoted $\mathcal{C} = \{\lambda(\theta), \theta \in \Theta \subset \mathbb{R}^d\}$, where θ is the model parameter. We want to determine if $\mathcal{C}$ is a relevant model for the observed data $T_1, \dots, T_n$. A goodness-of-fit test is a statistical test of $H_0 : \lambda \in \mathcal{C}$ vs $H_1 : \lambda \notin \mathcal{C}$. Usually, the test procedure consists in rejecting the null hypothesis of a good fit if some quantity, the test statistic, is higher than a critical value. This critical value is the quantile of either the exact or the asymptotic distribution of the statistic under H_0 . So the problem is first to find test statistics which express the gap between the data and the model, and second to determine the distribution of the statistics under H_0 .

In the following, we propose two families of goodness-of-fit tests for imperfect maintenance models that are based respectively on martingale residuals or probability integral transforms. For each test, the quantiles of the test statistic distribution under H_0 are computed with parametric bootstrap methods. The methodology is general and can be applied to a wide range of imperfect maintenance models.

3. Parametric bootstrap goodness-of-fit tests

3.1. Tests based on martingale residuals

Let $\Lambda = (\Lambda_t)_{t \geq 0}$ denote the cumulative intensity of the process N , such that $\Lambda_t = \int_0^t \lambda_s ds$, for $t \geq 0$. The process $M = (M_t)_{t \geq 0}$ defined by $M = N - \Lambda$ is a mean zero martingale. Then N is close to Λ in the sense that the expectation of their difference is null. In our setting, the intensity has a parametric form and is denoted $\lambda(\theta) = (\lambda_t(\theta))_{t \geq 0}$, for $\theta \in \Theta \subset \mathbb{R}^d$. The cumulative intensity is $\Lambda(\theta)$ and the corresponding martingale is $M(\theta) = N - \Lambda(\theta)$.

In practice, the parameter θ is unknown and the cumulative intensity is estimated from the n first failure times $T_1, \dots, T_n$. Let $\hat{\theta}$ denote the maximum likelihood estimator of θ , computed by maximizing (9). The random variables $\hat{M}_1, \dots, \hat{M}_n$ defined by

$$\hat{M}_i = N_{T_i} - \Lambda_{T_i}(\hat{\theta}) = i - \Lambda_{T_i}(\hat{\theta}), \quad i \in \{1, \dots, n\},$$

are called martingale residuals. The martingale property is lost on the $\hat{M}_i$ but N is still expected to be close to $\Lambda(\hat{\theta})$.

Then, the first family of goodness-of-fit tests is constructed on measures of discrepancies between N and $\Lambda(\hat{\theta})$. The tests reject the assumption that the model is valid if the two processes are too far apart. We have considered three usual test statistics based on the martingale residuals.

The first one is a Kolmogorov-Smirnov (KS) type statistic:

$$KS_m(\hat{\theta}) = \sup_{i=1, \dots, n} |\hat{M}_i| = \sup_{i=1, \dots, n} |i - \Lambda_{T_i}(\hat{\theta})|. \quad (10)$$

The second statistic is of Cramér-von Mises (CvM) type:

$$CvM_m(\hat{\theta}) = \int_0^{T_n} (N_t - \Lambda_t(\hat{\theta}))^2 d\Lambda_t(\hat{\theta}) = -\frac{1}{3} \sum_{i=1}^n \left\{ (i - 1 - \Lambda_{T_i}(\hat{\theta}))^3 - (i - 1 - \Lambda_{T_{i-1}}(\hat{\theta}))^3 \right\}. \quad (11)$$

We also propose a statistic of Anderson-Darling (AD) type:

$$\begin{aligned}
 AD_m(\hat{\theta}) &= \int_0^{T_n} \frac{(N_t - \Lambda_t(\hat{\theta}))^2}{\Lambda_t(\hat{\theta}) (n+1 - \Lambda_t(\hat{\theta}))} d\Lambda_t(\hat{\theta}) \\
 &= \frac{1}{n+1} \sum_{i=2}^n \left\{ (i-1)^2 \log \left(\frac{\Lambda_{T_i}(\hat{\theta})}{\Lambda_{T_{i-1}}(\hat{\theta})} \right) - (n+2-i)^2 \log \left(\frac{n+1 - \Lambda_{T_i}(\hat{\theta})}{n+1 - \Lambda_{T_{i-1}}(\hat{\theta})} \right) \right\} \\
 &\quad + (n+1) \log \left(1 - \frac{\Lambda_{T_1}(\hat{\theta})}{n+1} \right) - n.
 \end{aligned} \tag{12}$$

The distributions of the test statistics under the null hypothesis are not standard distributions and, in addition, they may depend on the parameters. Therefore, we will have to evaluate their quantiles by parametric bootstrap.

3.2. Tests based on probability integral transform

The second class of tests is based on the random variables $\Lambda_{T_{i+1}}(\theta) - \Lambda_{T_i}(\theta)$, for $i = 0, \dots, n-1$. Under H_0 , these variables are i.i.d. with standard exponential distribution. Let us transform these variables into uniform ones. For $i = 0, \dots, n-1$, let $S(\cdot | \mathbf{T}_i; \theta)$ denote the reliability function of the inter-failure time $T_{i+1} - T_i$ conditionally to $\mathbf{T}_i = (T_1, T_2, \dots, T_i)$:

$$S(s | \mathbf{T}_i; \theta) := P(T_{i+1} - T_i > s | \mathbf{T}_i; \theta) = \exp(-\Lambda_{T_i+s}(\theta) + \Lambda_{T_i}(\theta)), \quad \text{for } s \geq 0. \tag{13}$$

Let us define the variables $U_i(\theta) = S(T_{i+1} - T_i | \mathbf{T}_i; \theta)$ for $i = 0, \dots, n-1$. Under the null hypothesis, $H_0 : \lambda \in \mathcal{C}$, the U_i 's are i.i.d. with standard uniform distribution $\mathcal{U}[0, 1]$. Such a transformation of the inter-failure times is usually called Probability Integral Transform (PIT) and consists in applying a cumulative distribution function (c.d.f.) to a random variable. In our case, the c.d.f. is conditional on the past and the transformation is the Rosenblatt's one, then we call it conditional PIT.

The second class of goodness-of-fit tests that we propose in this article is based on the conditional PIT of the inter-failure times. We can expect that the uniformity will no longer hold in case of model misspecification. Therefore, we can test the goodness of fit of an imperfect maintenance model by testing that the transformed inter-failure times have a uniform distribution. In applications, θ is estimated and we consider the KS, CvM and AD test statistics for testing the uniformity of $U_0(\hat{\theta}), \dots, U_{n-1}(\hat{\theta})$. The test statistics are respectively

$$KS_u(\hat{\theta}) = \sqrt{n} \sup_{x \in [0,1]} |F_{n,S}(x) - x| = \sqrt{n} \max \left\{ \max_{i=1, \dots, n} \left(\frac{i}{n} - U_{(i-1)}(\hat{\theta}) \right), \max_{i=1, \dots, n} \left(U_{(i-1)}(\hat{\theta}) - \frac{i-1}{n} \right) \right\}, \tag{14}$$

$$CvM_u(\hat{\theta}) = n \int_0^1 (F_{n,S}(x) - x)^2 dx = \sum_{i=1}^n \left(U_{(i-1)}(\hat{\theta}) - \frac{2i-1}{2n} \right)^2 + \frac{1}{12n}, \tag{15}$$

$$AD_u(\hat{\theta}) = n \int_0^1 \frac{(F_{n,S}(x) - x)^2}{x(1-x)} dx = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) \{ \log(U_{(i-1)}(\hat{\theta})) + \log(1 - U_{(n-i)}(\hat{\theta})) \}, \tag{16}$$

where $F_{n,S}$ is the empirical c.d.f. of the random variables $U_i(\hat{\theta})$, and $U_{(0)}(\hat{\theta}) \leq U_{(1)}(\hat{\theta}) \leq \dots \leq U_{(n-1)}(\hat{\theta})$ are the ordered $U_i(\hat{\theta})$, $i = 0, 1, \dots, n-1$.

Liu et al (2012) proposed to perform a goodness-of-fit test by comparing the value of $KS_u(\hat{\theta})$ to usual critical values for testing the uniformity of a sample, but the $U_i(\hat{\theta})$ are neither independent nor uniformly distributed, so this method is not exact. Lindqvist et al (2003) used another transform of the $\Lambda_{T_{i+1}}(\theta) - \Lambda_{T_i}(\theta)$, so their method is close to ours. We have performed a numerical experiment which seems to indicate that the distributions of these statistics depend on the model parameters. So it is necessary to use a parametric bootstrap approach to perform the tests.

3.3. Construction of the tests by parametric bootstrap

Our test procedure is adapted from the article of Stute et al (1993), in which parametric bootstrap goodness-of-fit tests are built for i.i.d. random variables. Let $Z(\hat{\theta})$ denote the generic test statistic. In this article, $Z(\hat{\theta})$ will be one of the statistics $KS_m(\hat{\theta}), CvM_m(\hat{\theta}), AD_m(\hat{\theta}), KS_u(\hat{\theta}), CvM_u(\hat{\theta})$ or $AD_u(\hat{\theta})$, but other statistics could be

used. Under H_0 , $Z(\hat{\theta})$ is computed from the dataset $T_1, \dots, T_n$ generated from a point process with intensity $\lambda(\theta)$. We would like to obtain i.i.d. replications of $Z(\hat{\theta})$ to compute the empirical quantiles of the statistic distributions. However, θ is unknown. Then, θ is estimated by $\hat{\theta}$ and we simulate i.i.d. replications $T_1^*, \dots, T_n^*$ from a point process with intensity $\lambda(\hat{\theta})$. For each replication, the maximum likelihood estimator $\hat{\theta}^*$ and the test statistic $Z^*(\hat{\theta}^*)$ can be computed. The closeness of θ and $\hat{\theta}$ makes us expect a very low difference between the empirical quantiles of the distributions of $Z(\hat{\theta})$ and $Z^*(\hat{\theta}^*)$.

The general procedure for applying the test is described in the following algorithm:

1. Compute the MLE $\hat{\theta}$ of θ in the class of models $\mathcal{C}$ and compute the statistic $Z(\hat{\theta})$ on the dataset $T_1, \dots, T_n$.
2. For $i = 1$ until L ,
 - a. Generate $T_{1,i}^*, T_{2,i}^*, \dots, T_{n,i}^*$ under the model of intensity $\lambda(\hat{\theta}) \in \mathcal{C}$.
 - b. Compute $\hat{\theta}_i^*$ the MLE of $\hat{\theta}$ from $T_{1,i}^*, \dots, T_{n,i}^*$ in the model $\mathcal{C}$.
 - c. Compute the statistic $Z_i^* = Z_i^*(\hat{\theta}_i^*)$ from $T_{1,i}^*, \dots, T_{n,i}^*$ and $\hat{\theta}_i^*$.
3. The hypothesis H_0 is rejected at significance level α if $Z(\hat{\theta})$ is higher than the empirical quantile of order $1 - \alpha$ of $Z_1^*, \dots, Z_L^*$.

The simulations and the computations of Step 2 are the parametric bootstrap part of the algorithm.

4. Simulation study

To assess their quality, the tests are performed on a huge number of simulated datasets. The power of a test is estimated by the percentage of rejection of the null hypothesis.

The test level is set to $\alpha = 0.05$. We have considered three null hypotheses, all with PLP initial intensities. Under the first null hypothesis, the maintenance model is ARA_∞ , under the second one, the maintenance model is ARA_1 and under the third one, the maintenance model is QR. These models are denoted respectively ARA_∞ -PLP, ARA_1 -PLP and QR-PLP.

The test powers are evaluated by simulating models which are not the model tested in H_0 . For the maintenance effect, we chose ARA_1 , ARA_∞ , BP, QR and EGP models. For ARA models, we have set $p \in \{0.2, 0.8\}$, representing a low and a strong maintenance efficiency. Similarly, $p \in \{0.2, 0.8\}$ for the BP model. For the QR model, we have set $q \in \{0.8, 0.9, 0.95\}$. Indeed, when q is close to 1, the counting process looks like a renewal process. When q is too small, the inter-arrival times are too small and a phenomenon of explosion appears. For the EGP model, we have set $q \in \{0.8, 0.9, 0.95\}$ and $b_i = i - 1$ or $b_i = \sqrt{i - 1}$ for $i \in \{1, \dots, n\}$. For the initial intensity, we first chose a PLP with scale parameter $a = 0.05$ and shape parameter $b \in \{1.5, 2, 2.5, 3\}$, corresponding to mild to strong wear. We have also considered the LLP initial intensity with $a = -5$ and $b \in \{0.005, 0.01, 0.05, 0.1\}$. For each of these models, $M = 1500$ datasets are simulated, each composed of $n = 30$ failures.

The parameters are estimated by likelihood maximization under the constraints that $b \geq 1$ and $\rho \in [0, 1]$. The constraints ensure that the system is wearing and the repair is efficient. For the evaluation of the quantiles of the statistic distributions, $L = 1000$ bootstrap samples are generated.

During the talk, some simulation results will be showed. For small sample sizes ($n \leq 30$), the power of the tests is quite poor. But it increases significantly for large data sets ($n \geq 100$).

The main conclusion of this simulation study is that we recommend to test the goodness of fit of the ARA_∞ -PLP, the ARA_1 -PLP or the QR-PLP models both with the AD_m test built on martingale residuals and the AD_u test based on the conditional PIT. It is very difficult to distinguish the ARA_∞ and the BP models because both are close to a renewal process. The ARA_1 -PLP and QR-PLP models seem to be very flexible as they are able to approach a large range of alternative models. Therefore, these models are hardly ever rejected.

The goodness-of-fit tests of the ARA -PLP models are not powerful when data are simulated with an initial intensity of type LLP. We conclude that the proposed tests are more able to detect a discrepancy in the repair effect than in the shape of the intrinsic wear. This conclusion does not hold for QR models because the test powers are higher when the initial intensity is misspecified.

5. Applications

We have applied the goodness-of-fit tests for the ARA_{∞} -PLP, ARA_1 -PLP and QR-PLP models to several real datasets.

The first dataset concerns 4 identical and independent systems in 4 different EDF coal-fired power stations, denoted S_1 , S_2 , S_3 and S_4 . The dates of maintenance of the 4 systems are collected over a time-period of 9 years. Respectively, 22, 23, 16 and 31 failures have been observed. For S_1 and S_2 , none of the 3 models is rejected. With the AIC criterion, we select the QR model for S_1 and the ARA_{∞} model for S_2 . For S_3 and S_4 , all models are rejected. This may be because these data exhibit long time periods without any failure: this phenomenon can not be captured by our models.

The second dataset is composed of the maintenance times of a photocopier (Murthy et al, 2003). During four and a half years after being put in service, the system has been repaired $n = 42$ times. Doyen (2010) considered an ARA_1 -PLP model on this dataset. Our tests reject the ARA_{∞} -PLP model, but not the two others. So the choice of the ARA_1 -PLP model can be considered as valid. However, the AIC of the ARA_1 model is higher than the one of the QR model so we would recommend to use this latter model.

The last considered example is the AMC ambassador dataset (Ahn et al, 1998), which is composed of $n = 18$ failure times of an AMC Ambassador car owned by the Ohio state government. All tests do not reject the 3 considered models but the p-values of the tests of the ARA_{∞} model are lower than those of the ARA_1 and QR models. The AIC leads to select the ARA_1 -PLP model, as in Corset et al (2012).

Acknowledgement

This work was supported by the French National Agency of Research (ANR), project AMMSI, number ANR-2011-BS01-021.

References

- Ahn CW, Chae KC, Clark GM (1998) Estimating parameters of the Power-Law process with two measures of time. *Journal of Quality Technology*, 30, 127-132.
- Bordes L, Mercier S (2013) Extended geometric processes: semiparametric estimation and application to reliability. *Journal of the Iranian Statistical Society*, 12, 1-34.
- Brown M, Proschan F (1983) Imperfect repair. *Journal of Applied Probability*, 20, 851-859.
- Corset F, Doyen L, Gaudoin O (2012) Bayesian analysis of ARA imperfect repair models. *Communications in Statistics, Theory and Methods*, 41, 3915-3941.
- Doyen L (2010) Asymptotic properties of imperfect repair models and estimation of repair efficiency. *Naval Research Logistics*, 57, 296-307.
- Doyen L, Gaudoin O (2004) Classes of imperfect repair models based on reduction of failure intensity or virtual age. *Reliability Engineering and System Safety*, 84 (1), 45-56.
- Kijima M (1989) Some results for repairable systems with general repair. *Journal of Applied Probability*, 26, 89-102.
- Lam Y, Zhang YL (2003) A geometric-process maintenance model for a deteriorating system under a random environment. *IEEE Transactions on Reliability*, 52 (1), 83-88.
- Lindqvist B, Elvebakk G, Heggland K (2003) The trend-renewal process for statistical analysis of repairable systems. *Technometrics*, 45, 31-44.
- Liu Y, Huang HZ, Zhang X (2012) A data-driven approach for selecting imperfect maintenance models. *IEEE Transactions on Reliability*, 61, 101-112.
- Murthy DNP, Xie M, Jiang R (2003) *Weibull models*. Wiley.
- Stute W, Manteiga WG, Quindmil MP (1993) Bootstrap based goodness-of-fit tests. *Metrika*, 40, 243-256.
- Wang H, Pham H (1996) A quasi-renewal process and its applications in imperfect maintenance. *International Journal of Systems Science*, 27 (10), 1055-1062.

Residual life estimation based on multi-phase nonlinear wiener process

Hao Chen* and Bo Guo**

*College of Information System and Management, National University of Defense Technology, Changsha, China +
chenhao_nudt@163.com

** College of Information System and Management, National University of Defense Technology, Changsha, China +
boguo@nudt.edu.cn

Abstract. Residual life (RL) estimation based on multi-phase nonlinear Wiener process was studied in this paper, which is significant for complicated products with small samples. Firstly, nonlinear Wiener model with random parameter was introduced and multi-phase nonlinear Wiener model was proposed to model degradation process of products that were nonlinear and separated into different phases. Then the multi-phase RL probability density function based on the presented model was derived approximately in a closed form and parameters estimation was achieved with the method of maximum likelihood estimation (MLE). Finally, the method was applied to estimate the RL of high voltage plus capacitor. Compared with the other three different models by log-likelihood function (Log-LF) and Akaike information criterion (AIC), the results show that the proposed degradation model can capture degradation process of high voltage plus capacitors in a better way and provide a more reliable result.

1. Introduction

With rapid development of science and technology, designs for products become more and more precise, and correspondingly their failure modes and failure mechanisms are more and more complicated. Therefore, it may be the case that the degradation process can be divided into multi-phase and each phase has different features. Besides, the degradation path of performance characteristic may be nonlinear as well as it would be difficult to be linearized by transformation. The issue of multi-phase nonlinear Wiener process can link them together, which plays an important role in reliability analysis for small-sample, long-life and highly reliable products. Hence, more efforts should be given to this issue. This lead to the main purpose of this paper, to investigate the method of residual life (RL) estimation based on multi-phase nonlinear Wiener process.

With the performance degradation, the products' reliability declines along. As we known, the degradation process can be monotonic or non-monotonic. For the non-monotonic degradation process, stochastic Wiener model performs a better description. Many scholars have made their contribution to the study about Wiener process while most literatures are studied under the assumption that the degradation process is single-phase linear Wiener process. Peng (2010) proposed Bayes estimation method based on linear Wiener process and analyzed the bivariate degradation of products based on linear Wiener process. Si (2013) developed a Wiener-process-based degradation model, which used a recursive filter to update the drift coefficient in the Wiener process and used the expectation maximization (EM) algorithm to update all other parameters. With a log-transformation to luminance, Tang (2008) used a Wiener process with a linear drift to implement reliability estimation. By a time-transformation, Wang (2010) modeled the performance degradation of the beam bridge with a linear drift-based wiener process.

However, the assumption may not be realistic for some practical cases. On the one hand, multi-phase degradation process is existing. Feng (2012) proposed multi-phase Wiener degradation model to achieve the storage life prediction for a high-performance capacitor. Wang (2012) proposed a method about real-time reliability evaluation for the equipment whose degradation process was separated into different phases and the method was applied to liquid coupling device (LCD). Li (2009) analyzed the multi-phase degradation process with random change-point, built up a linear Wiener process model with random change-point and applied the model to the metalized film capacitors. More literature about multi-phase degradation process referred to Wang (2014) and Bae and Kvam (2007). On the other hand, nonlinear degradation process is also existing. In practical, not all nonlinearities within the degradation process can be properly linearized. Wang (2013) presented an adaptive method of residual life estimation based on a generalized Wiener degradation process, which took the nonlinearity. Si (2012) presented a nonlinear Wiener process-based degradation model to capture the nonlinear degradation without data transformation. More literature about nonlinear

degradation process referred to Wang (2010) and Tseng and Peng (2007). This paper takes high voltage plus capacitors as example and proposes multi-phase nonlinear Wiener process. Based on the proposed model, the probability density function (PDF) of RL is derived in a closed-form. Besides, the parameters are estimated with the method of Maximum likelihood estimation MLE. Last but not least, the proposed model is compared with the other three model in log-likelihood function (Log-LF) and Akaike information criterion (AIC). The results both confirm the validity and advantage of the proposed model.

The rest of this paper is organized as follows. In Section 2, the multi-phase nonlinear Wiener degradation model is presented. In Section 3, the PDF of RL is derived and the method of RL estimation is developed. Section 4 deals with inference for the unknown parameters. In Section 5, the example of high voltage plus capacitors is studied to illustrate the superiority of the proposed model through comparison with other models. Lastly, some conclusions are made in section 6.

2. Multi-phase nonlinear Wiener model

An example of multi-phase nonlinear degradation process is shown in Fig. 1. According to engineering experience, capacitance is an important performance characteristic of capacitors because it will directly determine capacitors' operational performance. Ye (2007) did a research about high voltage plus capacitors and found out that the capacitance variability varied throughout the storage time obviously and had an apparent characteristic of multi-phase. As depicted in Fig.1, in the beginning of storage time, the capacitance vibrates firstly but keep steady as a whole, while in the medium of storage time the capacitance increases sharply. Compared with the first two variation, the capacitance increases slowly in the later storage time.

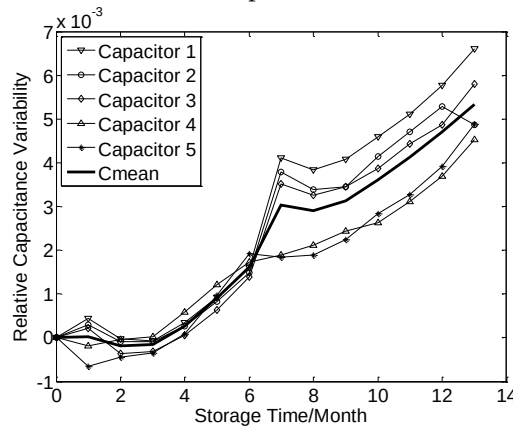


Fig.1 Relative capacitance variability with storage time

We let $X(t)$ denote the relative capacitance variability at time t , that is $X(t) = (C(t) - C(0)) / C(0)$, where $C(t)$ is the actual capacitance after storing t month and $C(0)$ is the initial capacitance.

For storage degradation analysis, if the relative capacitance variability $X(t)$ exceeds a pre-specified failure threshold ω , the capacitor is announced to be failed. Here the failure threshold ω is defined as 5%. In order to describe the degradation process of high voltage plus capacitor in a better way, in the proposed multi-phase nonlinear degradation model, we make following assumption.

Assumption 1: Ye (2007) pointed out that the degradation of high voltage plus capacitor had a characteristic of three phases with different features, however, the respective storage time of the first two phases were too short, as a result, degradation data measured was limited, which is going against modeling for each phase precisely. Here we consider to treat the first phase and the second phase as a whole phase. So we assumed that the degradation process consists of two different phases.

Assumption 2: A nonlinear Wiener degradation model is adopted to model cumulative degradation for the two phases. According to Wang (2014), utilizing the generalized Wiener degradation model, $X(t)$ can be described as

$$X(t) = \mu \int_0^t \lambda(t; \theta) dt + \sigma B(\tau(t; \gamma)). \quad (1)$$

Where μ is drifting coefficient, σ is diffusion coefficient ($\sigma > 0$) and $B(\cdot)$ is the standard Brownian motion. Both $\lambda(t; \theta)$ and $\tau(t; \gamma)$ are non-decreasing functions. Here let $\Lambda(t; \theta) = \int_0^t \lambda(t; \theta) dt$, then $X(t)$ can be represented as

$$X(t) = \mu\Lambda(t; \theta) + \sigma B(\tau(t; \gamma)). \quad (2)$$

In order to reflect the product-to-product variability, it is assumed that the drifting coefficient μ is a normally distributed variable with mean μ_0 and variance σ_0^2 . Besides, diffusion coefficient σ is assumed as a fixed value. Let $\Theta = \{\theta, \gamma, \mu_0, \sigma_0, \sigma\}$, $X_1(t; \Theta_1)$ and $X_2(t; \Theta_2)$ denote the nonlinear Wiener degradation model given the parameters for the two phases, respectively. Then, $X(t)$ can be denoted as

$$X(t) = \begin{cases} X_1(t; \Theta_1) & 0 < t \leq \tau \\ X_2(t - \tau; \Theta_2) & t > \tau \end{cases}. \quad (3)$$

Where τ denotes the change-point time between the first phase and second phase, and for simplicity τ is assumed as a fixed value, which is determined by Ye (2007), $\tau = 7$.

Assumption 3: On the one hand, empirical studies show that the mean degradation over time is in proportion to time with a power law, on the other hand, the characteristic of mean deterioration is accord with power function. Hence, it is assumed that $\Lambda(t; \theta) = t^\theta$, $\tau(t; \gamma) = t^\gamma$.

3. Residual life estimation

Since the multi-phase nonlinear Wiener degradation model is complicated and it is difficult to derive an exact expression of residual life distribution, an approximate failure time distribution will be adopted.

For convenience, let $\omega_h = \omega - X(t_h)$, $X(t_h)$ represents the capacitance variability at time t_h . Denoted the RL at time t_h as L , the RL can be regarded as the first passage time of $X(t_h + l) - X(t_h)$ exceeding ω_h , that is

$$L = \inf \{l_h : X(t_h + l) - X(t_h) \geq \omega_h \mid X(t_h)\}. \quad (4)$$

Where $X(t_h + l) - X(t_h) = \mu\Delta\Lambda(t_h + l; \theta) + \sigma B(\Delta\tau(t_h + l; \gamma))$, $\Delta\Lambda(t_h + l; \theta) = \Lambda(t_h + l; \theta) - \Lambda(t_h; \theta)$, $\Delta\tau(t_h + l; \gamma) = \tau(t_h + l; \gamma) - \tau(t_h; \gamma)$. Let $s_h = \Delta\tau(t_h + l; \gamma)$, $Y_h(s_h) = X(t_h + l) - X(t_h)$, then $Y_h(s_h)$ can be transformed as follows $Y_h(s_h) = \mu\eta(s_h; \theta) + \sigma B(s_h)$.

Where $\eta(s_h; \theta) = \Delta\Lambda(t_h + \nu(s_h; \gamma); \theta)$ and $\nu(s_h; \gamma)$ is solved out from the inverse of $\Delta\tau(t_h + l; \gamma)$.

Denoted S_h as the first passage time of $Y_h(s_h)$ exceeding ω_h , S_h can be represented as $S_h = \inf \{s_h : Y_h(s_h) \geq \omega_h\}$. Considering the relationship between L and S_h , it can be derived that

$$g(l \mid \mu) = g_s(s_h \mid \mu) \frac{d\Delta\tau(t_h + l; \gamma)}{dl} \quad (5)$$

Where $g(l \mid \mu)$ is the conditional PDF of RL and $g_s(s_h \mid \mu)$ is the conditional PDF of S_h . In other word, the key to derive the expression of $g(l \mid \mu)$ is the determination of $g_s(s_h \mid \mu)$.

To determine $g_s(s_h \mid \mu)$, the work of Si (2012) is introduced. They has done research on the model $X(t) = \mu\Lambda(t; \theta) + \sigma B(t)$, which was adopted to model the degradation for lasers, gyroscopes and Fatigue crack. Under a mild assumption, the approximated conditional PDF of the failure time was showed as

$$f(t \mid \mu) = \frac{1}{\sqrt{2\pi t}} \left(\frac{(D - \mu\Lambda(t; \theta)) / \sigma}{t} + \frac{\lambda(t; \theta)}{\sigma} \right) \cdot \exp \left(- \frac{((D - \mu\Lambda(t; \theta)) / \sigma)^2}{2t} \right) \quad (6)$$

Combining $Y_h(s_h)$ and (8), the conditional PDF of S_h can be represented as

$$g_s(s_h \mid \mu) = \frac{1}{\sqrt{2\pi s_h}} \cdot \exp \left(- \frac{((D_h - \mu\eta(s_h; \theta)) / \sigma)^2}{2s_h} \right) \cdot \left(\frac{(D_h - \mu\eta(s_h; \theta)) / \sigma}{s_h} + \frac{\mu d\eta(s_h; \theta)}{\sigma ds_h} \right). \quad (7)$$

Therefore, according to (7), the conditional PDF of RL can be approximated as

$$g(l \mid \mu) = \frac{1}{\sqrt{2\pi s_h}} \cdot \exp \left(- \frac{(D_Y(s_h))^2}{2s_h} \right) \cdot \left(\frac{D_Y(s_h)}{s_h} + \frac{\kappa(s_h; \theta)}{\sigma_0} \right) \cdot \frac{ds_h}{dl}. \quad (8)$$

Besides, when the uncertainty of μ is taken into consideration, then the PDF of RL can formulated as

$$g(l) \cong \int_{-\infty}^{+\infty} g(l \mid \mu) \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp \left(- \frac{(\mu - \mu_0)^2}{2\sigma_0^2} \right) d\mu. \quad (9)$$

Since the PDF $g(x)$ should satisfy $\int_{\Phi} g(x) = 1$. Then the PDF of RL can be re-expressed as

$$f(l) = \frac{1}{\int_0^{\infty} g(l) dl} \cdot \frac{1}{s_h \sqrt{2\pi B_h}} \cdot \exp\left(-\frac{(\omega_h - \mu\eta(s_h; \theta))^2}{2B_h}\right) \cdot \left(\omega_h - C_h \cdot \frac{\omega_h \eta(s_h; \theta) \sigma_0^2 + \mu \sigma^2 s_h}{B_h}\right) \cdot \frac{ds_h}{dl} \quad (10)$$

With multi-phase nonlinear Wiener model proposed above, the PDF of RL in different phase will be different. As discussed above, in the first phase the degradation can be modeled by $X_1(t; \Theta_1)$, and the PDF of RL can be divided into two parts: when $0 < RL < \tau - t_h$, it can be captured by $f_1(l; \Theta_1, \omega)$, when $RL > \tau - t_h$, it can be described by $f_2(l; \Theta_2, \omega)$. As the high voltage plus capacitor degraded to the time τ , the degradation has to be re-modeled by $X_2(t; \Theta_2)$ and the PDF of RL is changed to $f_2(l; \Theta_2, \omega)$, directly. According to the above analysis, following PDF of RL are adopted to residual life estimation for the proposed degradation model.

$$f(l) = \begin{cases} f_1(l; \Theta_1, \omega) & 0 < t_h \leq \tau, 0 < RL < \tau - t_h \\ f_2(l - \tau; \Theta_2, \omega - X(\tau)) & 0 < t_h \leq \tau, RL > \tau - t_h \\ f_2(l - \tau; \Theta_2, \omega - X(\tau)) & t_h > \tau \end{cases} \quad (11)$$

4. Parameter estimation

According to proposed the multi-phase nonlinear Wiener Process, we need to estimate the parameters of two phases, $\Theta_i = \{\theta_i, \gamma_i, \mu_{0i}, \sigma_{0i}, \sigma_i\}$, $i=1, 2$. To achieve the parameters estimation, we assume that there are n capacitors to be tested and all the capacitors are measured at the same time and the total measurement times m for the capacitors are identical. We denoted the capacitance variability measured for the i th capacitor at the time t_h as $X_i(t_h)$, $i=1, \dots, n, h=1, \dots, m$. The capacitance variability measured in different phase for the i th capacitors can be represented as

$$\begin{cases} \mathbf{X}_{i1} = (X_i(t_1), X_i(t_2), \dots, X_i(t_{\tau_1})) \\ \mathbf{X}_{i2} = (X_i(t_{\tau_{k-1}+1}), X_i(t_{\tau_{k-1}+2}), \dots, X_i(t_{m_i})) \end{cases} \quad (12)$$

Since the capacitors are measured at the same time, we denote $\Lambda_1 = (\Lambda(t_1; \theta_1), \Lambda(t_2; \theta_1), \dots, \Lambda(t_{\tau_1}; \theta_1))$ for the first phase. Because drifting coefficient μ is a normally distributed variable with mean μ_0 and variance σ_0^2 , clearly, $\mathbf{X}_{i1}$ is normally distributed with mean and covariance $\mu_1 \Lambda_1$ and $\Sigma_1 = \Omega_1 + \sigma_1^2 \Lambda_1 \Lambda_1'$, respectively. That is $\mathbf{X}_{i1} \sim N(\mu_1 \Lambda_1, \Sigma_1)$.

According to (16), the log-likelihood function of Θ_1 is

$$l(\Theta_1 | \mathbf{X}_1) = -\frac{n \cdot \tau_1 \cdot \ln(2\pi)}{2} - \frac{\ln(|\Sigma_1|)}{2} - \frac{1}{2} \sum_{i=1}^n (\mathbf{X}_{i1} - \mu_1 \Lambda_1) \Sigma_1^{-1} (\mathbf{X}_{i1} - \mu_1 \Lambda_1)' \quad (13)$$

Where $\Theta_1 = \{\theta_1, \gamma_1, \mu_{01}, \sigma_{01}, \sigma_1\}$, $\mathbf{X}_1 = (\mathbf{X}_{11}, \dots, \mathbf{X}_{n1})$.

By differentiating $l(\Theta_1 | \mathbf{X}_1)$ in (13) with respect to μ_{01} and σ_{01} , we obtain

$$\hat{\mu}_{01} = \frac{\sum_{i=1}^n \Lambda_1 \Omega_1^{-1} \mathbf{X}_{i1}'}{m \Lambda_1 \Omega_1^{-1} \Lambda_1'}, \hat{\sigma}_{01} = \left\{ \frac{1}{m (\Lambda_1 \Omega_1^{-1} \Lambda_1')^2} \sum_{i=1}^m (\mathbf{X}_{i1} - \hat{\mu}_{01} \Lambda_1) \Omega_1^{-1} \Lambda_1' \cdot \Lambda_1 \Omega_1^{-1} \cdot (\mathbf{X}_{i1} - \hat{\mu}_{01} \Lambda_1)' - \frac{1}{\Lambda_1 \Omega_1^{-1} \Lambda_1'} \right\}^{1/2} \quad (14)$$

Maximizing the log-likelihood function $l(\theta_1, \gamma_1, \sigma_1 | \mathbf{X}_1, \hat{\mu}_{01}, \hat{\sigma}_{01})$, the maximization likelihood estimation of θ_1, γ_1 and σ_1 can be obtained. Then, the MLE of μ_{01} and σ_{01} can be obtained from (14).

Hence, the log-likelihood function of θ_1, γ_1 and σ_1 can be obtained by substitute $\hat{\mu}_{01}$ and $\hat{\sigma}_{01}$ for μ_{01} and σ_{01} in the (13), that is

$$l(\theta_1, \gamma_1, \sigma_1 | \mathbf{X}_1, \hat{\mu}_{01}, \hat{\sigma}_{01}) = -\frac{n \tau_1 \ln(2\pi)}{2} - \frac{n}{2} - \frac{n}{2} \ln |\Omega_1| - \frac{1}{2} \left(\sum_{i=1}^n \mathbf{X}_{i1} \Omega_1^{-1} \mathbf{X}_{i1}' - \frac{\sum_{i=1}^n (\Lambda_1 \Omega_1^{-1} \mathbf{X}_{i1})^2}{\Lambda_1 \Omega_1^{-1} \Lambda_1'} \right) - \frac{n}{2} \ln \left(\frac{\sum_{i=1}^m (\Lambda_1 \Omega_1^{-1} \mathbf{X}_{i1})^2}{m \Lambda_1 \Omega_1^{-1} \Lambda_1'} - \frac{\sum_{i=1}^m (\Lambda_1 \Omega_1^{-1} \mathbf{X}_{i1})^2}{m^2 \Lambda_1 \Omega_1^{-1} \Lambda_1'} \right) \quad (15)$$

Likewise, $\mathbf{X}_{i2}$ is normally distributed with mean $\mu_2\mathbf{\Lambda}_2$ and covariance $\mathbf{\Sigma}_2 = \mathbf{\Omega}_2 + \sigma_2^2\mathbf{\Lambda}_2\mathbf{\Lambda}_2'$, that is $\mathbf{X}_{i2} \sim N(\mu_2\mathbf{\Lambda}_2, \mathbf{\Sigma}_2), i=1, \dots, n$. The estimation of $\Theta_2 = \{\theta_2, \gamma_2, \mu_{02}, \sigma_{02}, \sigma_2\}$ can be derived as above.

5. Case study

Now we take high voltage plus capacitors for example, data used in Ye (2007), which have been shown in Fig. 1. High voltage plus capacitors have the characteristics of long-life, and high reliability, but during the storage test, capacitance will degrade over time which is an importance performance characteristic and determine capacitors' operational performance directly.

To compare the fitting of the proposed models, the Log-LF value and the Akaike information criterion (AIC) were both used, as well as the mean square errors (MSEs). The AIC can be calculated by

$$AIC = -2 \times [\max(\text{Log-LF})] + 2p. \quad (18)$$

Firstly, according to the proposed two-phase nonlinear Wiener model, parameters estimation of two-phase degradation can be listed as follows.

Table I Parameters estimation of two-phase degradation

Parameters	θ	γ	μ_0	σ_0	σ
Phase 1	4.6109	1.1803	3.6365×10^{-7}	1.6769×10^{-7}	3.3335×10^{-4}
Phase 2	1.3249	1.1770	1.7849×10^{-4}	2.4321×10^{-5}	2.2081×10^{-4}

To verify the validity of the proposed model, a simulation procedure is used to generate degradation paths of 50 products based on the proposed model with the parameters in Table I. The results are shown in Fig. 2. It can be observed that the two-phase degradation characteristic can be depicted as well as the feature of nonlinear. What's more, the model proposed capture the degradation well. Besides, we present the residual life probability density curves along time, which are shown in Fig. 3.

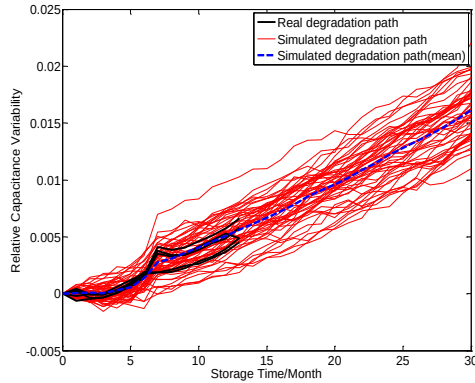


Fig. 2 Simulated degradation paths of 50 products

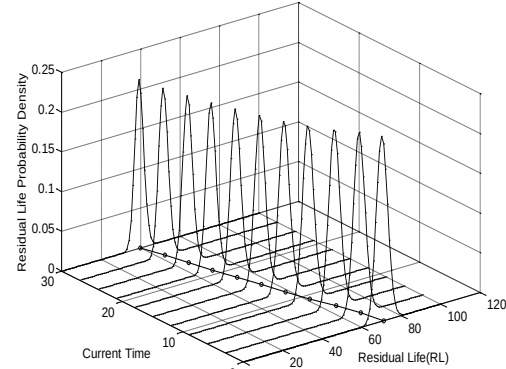


Fig. 3 Residual life probability density curve

To illustrate the superiority of the proposed model, we make further efforts on the comparison with other models, like Multi-phase linear Wiener model, Single-phase nonlinear Wiener model and Single-phase linear Wiener model. The parameters estimation for four different models are listed in Table II. The comparison of different models on failure time probability density and reliability are shown in Fig. 4.

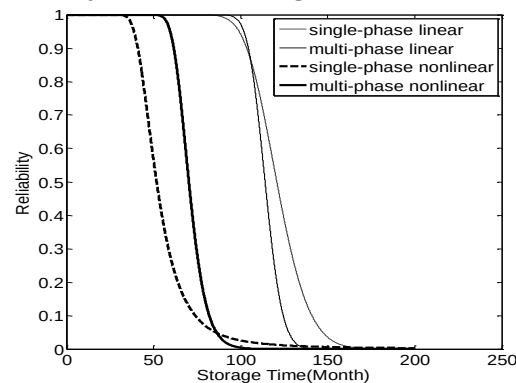
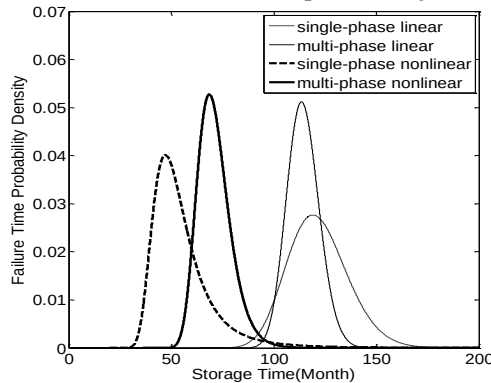


Fig. 4. (a) Comparison on failure time PDF

Fig. 4 (b) Comparison of different models on reliability curve

It can be seen from Table II that the Log-LF value of the proposed model is larger than the other three models. Besides, the AIC value of the proposed model is the smallest one among the four models.

Table II Parameters estimation of two-phase degradation

Model		θ	γ	μ	μ_0	σ_0	σ	Log-LF	AIC
Multi-phase nonlinear	Phase 1	4.6109	1.1803	/	3.6365×10^{-7}	1.6769×10^{-7}	3.3335×10^{-4}	419.0179	-818.0358
	Phase 2	1.3249	1.1770	/	1.7849×10^{-4}	2.4321×10^{-5}	2.2081×10^{-4}		
Multi-phase linear	Phase 1	1	1	-5.4067×10^{-5}	/	/	3.3905×10^{-4}	413.2649	-814.5298
	Phase 2	1	1	4.3266×10^{-4}	/	/	8.7865×10^{-4}		
	Phase 3	1	1	4.1091×10^{-4}	/	/	3.0119×10^{-4}		
Single-phase nonlinear	/	1.6048	1.2642	/	8.8475×10^{-5}	2.9728×10^{-5}	3.9864×10^{-4}	400.2388	-790.4776
Single-phase linear	/	1	1	4.1091×10^{-4}	/	/	5.4617×10^{-4}	363.9105	-723.8210

6. Conclusion

In this paper, under an assumption a multi-phase nonlinear Wiener model is proposed to depict the degradation process of high voltage plus capacitor. By means of simulation and comparison, the proposed model is proved to be valid and suitable for capturing the degradation process. Besides, RL of product is approximately estimated in a static way. It can be seen that the chosen of model is important, if not, the deviation of result will be large. The model in this paper proposed is limited. How to divide the degradation process for a nonlinear way and the chosen of nonlinear model should be studied more. What's more, the change-point is set in advance and the RL estimation is static. More efforts should be made for them.

References

- B.H. Peng, J.L. Zhou, Z.Q. Pan (2010) Bayesian method for reliability assessment of products with Wiener process degradation. *Systems Engineering-Theory & Practice* **30**, 543-549.
- Peng B H (2010) Research on Reliability Modeling Methods Based on Wiener Process. Ph.D. dissertation, National University of Defense Technology, Changsha, China.
- X. S. Si, W. B. Wang, C. H. Hu, M. Y. Chen and D. H. Zhou (2013) "A Wiener-process-based degradation model with a recursive filter algorithm for remaining useful life estimation," *Mechanical Systems and Signal Processing* **35**, 219-237.
- J. Tang and T. S. Su (2008) Estimating failure time distribution and its parameters based on intermediate data from a wiener degradation model. *Naval Research Logistics* **55**, 265-276.
- X. Wang (2010) Wiener processes with random effects for degradation data. *Journal of Multivariate Analysis* **101**, 340-351.
- J. Feng, Q. Sun and T. D. Jin (2012) Storage life prediction for a high-performance capacitor using multi-phase wiener degradation model. *Communications in Statistics-Simulation*. **41**, 1317-1335.
- X. L. Wang, B. Guo, Z. J. Cheng (2012) Real-time reliability evaluation of equipment based on separated-phase wiener-Einstein process. *Journal of Central South University* **2**, 534-540.
- Li M (2009) Research on the degradation-failure modeling and analysis methods of degradation path with random change-points. M.S. thesis National University of Defense Technology, Changsha, China.
- X. L. Wang, P. Jiang, B. Guo and Z. J. Cheng (2014) Real-time Reliability Evaluation for an Individual Product Based on Change-point Gamma and Wiener process. *Quality and Reliability Engineering Internation* **30**, 513-525.
- S. J. Bae and P. H. Kvam (2007) A change-point analysis for modeling incomplete burn-in for light displays. *IIE Transactions* **38**, 489-498.
- X. L. Wang, N. Balakrishnan, B. Guo (2013) Residual life estimation based on a generalized wiener process. *Reliability Engineering & System Safety* **124**, 13-23.
- X. S. Si, W. Wang, C. H. Hu, D. H. Zhou and M. G. Pecht (2012) Remaining useful life estimation based on a nonlinear diffusion degradation process. *IEEE Transaction on Reliability* **61**, 50-67.
- W. B. Wang, M. Carr, W. J. Xu, K. Kobbacy (2010) A model for residual life prediction based on Brownian motion with an adaptive drift. *Microelectronics Reliability* **51**, 285-293.
- S. T. Tseng, C. Y. Peng (2007) Stochastic diffusion modeling of degradation data. *Journal of Data Science* **5**, 315-333.
- H. F. Ye, Y. H. Yao and C. L. Yu (2007) Capacitance variability of high voltage pulse capacitors. *High Voltage Engineering* **33**, 37-41.

Building a multicriteria decision model for risk evaluation in gas pipelines

A. T de Almeida*, M. H. Alencar**and C. Medeiros***

Universidade Federal de Pernambuco, Recife, Pernambuco, Brazil.

*almeidatd@gmail.com

**marcelohazin@gmail.com

***cpmedeiros2 @gmail.com

Abstract.

This paper presents a multicriteria decision model in order to evaluate natural gas pipelines and establish actions with a view to mitigating the existing risks. Multi-Attribute Utility Theory (MAUT) is incorporated into the model, and account is taken of the decision maker's (DM's) preferences. Three risk dimensions are considered: Human, financial and environmental. To support the risk evaluation in sections of the pipeline, a global sensitivity analysis is conducted where a variation of model parameters according to probability distributions are observed by means of the Monte Carlo simulation. Finally, to make this analysis more complete, Kendall's correlation coefficient is used so as to measure associations between rankings arising from replications.

1. Introduction

A multidimensional risk analysis approach in pipelines has been discussed in the literature (Brito & de Almeida, 2009; Brito et al, 2010, Alencar & de Almeida (2010), Lins & de Almeida, (2012) showing the importance of incorporating different perspectives on how best to analyze risk. These models provide the decision-maker (DM) with a recommendation on prioritizing sections of a pipeline, based on the losses associated with each section. It is noteworthy that there is a growing concern about how best to consider prospects for risk analysis that go beyond examining the financial aspects (de Almeida et al 2015a).

Therefore, it is necessary to evaluate the sources of uncertainty associated with estimating risks in the pipeline, since different levels of uncertainty produce different recommendations for the DM. Therefore, a global sensitivity analysis model (GSA) is proposed in order to identify the parameters of sensitivity and robustness of the model (Saltelli et al 2008).

2. A multicriteria decision model to evaluate natural gas pipelines

Brito and de Almeida (2009) proposed a risk assessment model in natural gas pipelines that considered three perspectives of risk: financial (f), human (h) and environmental (e). Multi-Attribute Utility Theory (MAUT) is used to aggregate the multiple dimensions of risk by incorporating the DM's preferences, as recommended by Keeney & Raiffa (1976). The one-dimensional utilities for each risk perspective are obtained by considering the maximum losses.

Scenarios that are considered accidental are those that may occur due to two failure modes: a hole or a rupture of the pipeline. Moreover, the normal scenario (θ_N) is the one in which the pipeline is functioning properly under normal conditions and then for each scenario, the states of nature π_i (θ_{jm}) are obtained. The pipeline is divided into sections (a_i) and for each of these, the risk is calculated as Equation (1), where k is the constant scale, j is the failure mode, and m is the scenario analyzed.

$$\left\{ \pi_i(\theta_{jm}) \cdot \sum_j \sum_m \left\{ - \left[\begin{array}{l} + \int_h f(h | \theta_{jm, \alpha_i}) k_h u(h) dh \\ + \int_g f(g | \theta_{jm, \alpha_i}) k_g u(g) dg \\ + \int_e f(e | \theta_{jm, \alpha_i}) k_e u(e) de \end{array} \right] \right\} \right\} \{ (-1) \pi_i(\theta_N) \} \quad (1)$$

3. Sensitivity Analysis

As to the Sensitivity Analysis (SA), different approaches in the literature are considered which offer different forms of evaluation of the uncertainties inherent in a model, both those which only consider an evaluation criterion and those which consider multiple criteria (Alzbutas et al (2014), Brown et al (2013), Koornneef et al 2013).

In particular, a SA is conducted in order to jointly analyze the uncertainty of the parameters of this study, which is intended to identify sources of uncertainty and analyze them more efficiently. Aven (2016) states that the SA is associated with sources of uncertainty that may affect the risk and therefore deserve to be highlighted in the evaluation. Additionally, different ways of evaluating and determining the probabilities that may affect the risk category in accordance with Bedford et al (2015).

Therefore, we seek not only to analyze sensitive parameters, but also to analyze the robustness of the model (Tang et al 2016; Qian & Niffenegger, 2011) in order to provide the DM not only with a recommendation based on the calculated risk, but one based on the inherent uncertainties of the model. In other words, what is proposed is a set of priority sections at risk and a series of sections with an increased risk as a result of a specified level of uncertainty.

The Monte Carlo Simulation (MCS) is widely used in the literature in order to provide samples or to evaluate and quantify uncertainties (Shields et al, 2015; Sanseverino et al, 2014). Despite needing a high computational effort, MCS is rich and information that generates a map of the data produced in the simulation, thereby generating information on the model analyzed (Helton, 1993).

If the DM is to be provided with information, a parameter must be used that can analyze the outputs generated in the simulation and compare these with the original result. Therefore, the use of Kendall's τ correlation coefficient, which provides a measure of association between each simulation and the original rank. The values of this parameter ranges between -1 and 1, indicating perfect negative and positive correlation, respectively. In addition, a probability related to the value of τ obtained can be associated. The recommendation to the DM is made because of a known confidence level as in Siegel & Castellan (1988). Subsequent analysis of τ values identifies variations of this parameter in the simulations, thus promoting greater understanding of the behavior of the rankings (Daher & de Almeida, 2012).

The following shows a numerical application of the proposed model of SA, and gives the necessary steps for its implementation and subsequent analysis.

4. Numerical Application

A pipeline of approximately 20,000 meters is analyzed, which is divided into 9 sections. The maximum loss for each one is considered, and dimensional utilities are obtained from the DM's preferences (Figure 1). Then, the scale constants are obtained by the steps performed in accordance with Keeney & Raiffa. (1976) in a space of consequences as illustrated in Figure 2. Then, the risk values are calculated for each section and the set of sections is ranked in descending order of risk, thereby obtaining the original ranking (r_0), shown in Table 1.

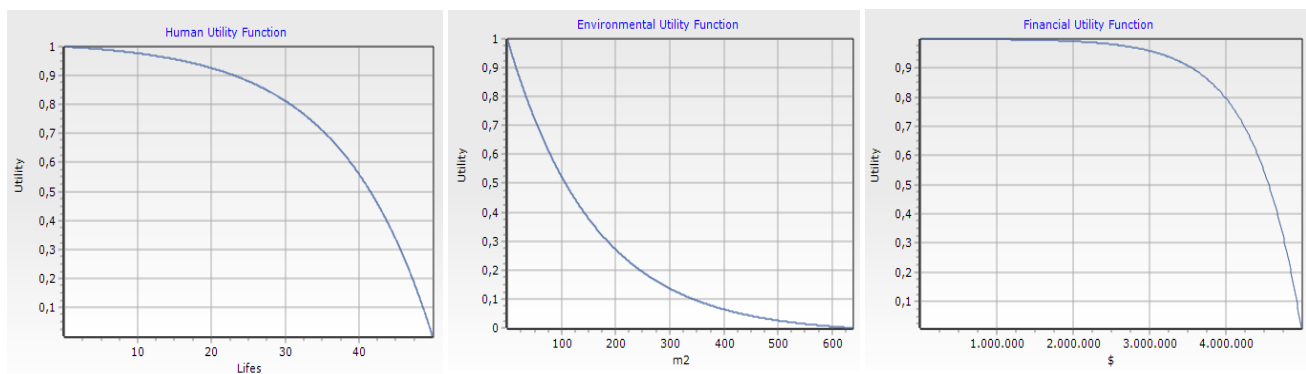


Figure 1. Unidimensional utilities for each risk perspective

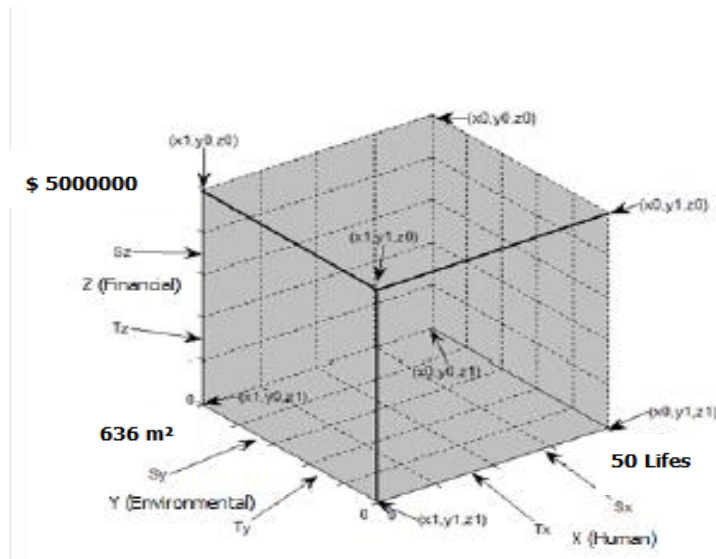


Figure 2. Space of consequences to obtain the scale constant (de Almeida et al. 2015b).

Table 1- Original ranking (r_0) and ratio

Position	Ranking	Absolute Difference (E-05)	Ratio
1	Section 1	4.0941	7.6779
2	Section 2	0.5332	0.2988
3	Section 4	1.7842	0.1786
4	Section 5	9.9870	1.6462
5	Section 9	6.0664	1.1120
6	Section 7	5.4554	0.4042
7	Section 8	13.4945	23.1819
8	Section 3	0.5821	
9	Section 6		

Four uncertainty ranges are considered: 5%, 10%, 15% and 20% for all parameters, including scale constants. For each of them, 10,000 simulations are obtained and their τ values are analyzed as shown in Table 2. The results show that a variation of 5% is quite robust at a confidence level of 95%, but for a range above 10%, larger variations are observed, as it is more sensitive, but still at a high confidence level (90%).

Table 2- τ values for each range simulated

Range	5%	10%	15%	20%
Minimum	0.1111	-0.3333	-0.3889	-0.5
Maximum	1.0000	1.0000	1.0000	1.0000
Mean	0.5994	0.4174	0.3741	0.3568
Median	0.6111	0.4444	0.3889	0.3889
Standard Deviation	0.1305	0.2218	0.233	0.2282
Mode	0.5556	0.4444	0.4444	0.3889
α	0.05	0.1	0.1	0.1

5. Conclusions

It was shown that the GSA identifies sources of uncertainty of a model, the MCS of which appears to be a simple and effective means in this assessment. In addition, assessing values based on τ enables the DM to

gain a greater understanding of the behavior of the variations in the pipeline sections, it being possible to present results to a known confidence level.

Is worth mentioning that the analysis of a set of parameters should be considered where this is needed, such as for a sensitive range. The robustness analyzed only deals with the data being studied and cannot be replicated to other pipelines, since the ranking will depend on the characteristics of each pipeline. However, it is considered that similar analysis may lead to results which enable recommendations to be made to the DM in order to mitigate the risk in pipelines.

References

- Alencar MH, de Almeida AT (2010) Assigning priorities to actions in a pipeline transporting hydrogen based on a multicriteria decision model. *International Journal of Hydrogen Energy* 35:3610-3619.
- Alzbutas R, Iešmantas T, Povilaitis M, Vitkutė J (2014) Risk and uncertainty analysis of gas pipeline failure and gas combustion consequence. *Stochastic environmental research and risk assessment* [1436-3240] vol:28 iss:6 pp:1431-1446.
- Aven T (2016) On the use of conservatism in risk assessments. *Reliability Engineering and System Safety* 146. 33–38
- Bedford T, Daneshkhah A, Wilson JK (2015) Approximate Uncertainty Modeling in Risk Analysis with Vine Copulas. *Risk Analysis*. DOI: 10.1111/risa.12471
- Brito AJ, de Almeida AT, Mota CMM (2010) A multicriteria model for risk sorting of natural gas pipelines based on ELECTRE TRI integrating utility theory. *European Journal of Operation Research* 200: 812 and 821.
- Brown S, Beck J, Mahgerefteh H, Fraga ES (2013) Global sensitivity analysis of the impact of impurities on CO₂ pipeline failure. *Reliability Engineering & Systems Safety* [0951-8320] vol:115 pp:43 -5.
- Daher SFD, de Almeida AT (2012) The use of ranking veto concept to mitigate the compensatory effects of additive aggregation in group decisions on a water utility automation investment. *Group Decision and Negotiation* v. 21, n. 2, p. 185-204.
- de Almeida AT, Ferreira RJP, Cavalcante AV (2015a) A review of the use of multicriteria and multi-objective models in maintenance and reliability. *IMA Journal of Management Mathematics* 26, 249–271. doi:10.1093/imaman/dpv010.
- de Almeida AT, Cavalcante CAV, Alencar MH, Ferreira RJP, de Almeida-Filho AT, Garcez TV (2015b) *Multicriteria and Multi-objective Models for Risk, Reliability and Maintenance Decision Analysis*. International Series in Operations Research & Management Science. Vol231. 387p. New York: Springer
- Helton JC (1993) Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering and System Safety* 42. 327-367.
- Keeney RL, Raiffa H (1976) *Decision with Multiple Objectives: Preferences and value trade-offs*. John Wiley & Sons.
- Koornneef J, Spruijt M, Molag M, Ramírez A, Turkenburg W, Faaij A (2010) Quantitative risk assessment of CO₂ transport by pipelines-A review of uncertainties and their impacts. *Journal of Hazardous Materials* [0304-3894] vol:177 iss:1-3 pp:12-27.
- Lins PHC, de Almeida AT (2012) Multidimensional risk analysis of hydrogen pipelines. *International Journal of Hydrogen Energy* 37, 13545-13554.
- Qian G, Niffenegger M (2011) Probabilistic fracture assessment of piping systems based on FITNET FFS procedure. *Nuclear Engineering and Design* 241, 714–722
- Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S (2008) *Global Sensitivity Analysis*. Wiley.
- Sanseverino CMR, Ramirez-Marquez JE (2014) Uncertainty propagation and sensitivity analysis in system reliability assessment via unscented transformation. *Reliability Engineering and System Safety* 132, 176-185.
- Shields MD, Teferra K, Hapij A, Daddazio RP (2015) Refined Stratified Sampling for efficient Monte Carlo-based uncertainty quantification. *Reliability Engineering and System Safety* 142, 310-325
- Siegel S, Castellan NJ (1988) *Nonparametric statistics for the behavioral sciences*, McGraw Hill, 1988.

Tang C, Zuo MJ, Xiao N (2016) An efficient method for evaluating the effect of input parameters on the integrity of safety systems. *Reliability Engineering and System Safety*145, 111-123.

Reliability assessment of buried pipelines under seismic action

Andrew Utomi Ebenuwa*, Kong Fah Tee**

*Department of Engineering Science, University of Greenwich, UK. Email address: ea106@gre.ac.uk

**Department of Engineering Science, University of Greenwich, UK. Email address: K.F.Tee@gre.ac.uk

Abstract: The assessment of failure probability of a buried steel water pipeline in seismic areas involves evaluating the uncertainties associated with the pipe performance due to earthquake actions. This paper presents a computational framework for implementing an advanced Monte Carlo simulation called Line Sampling (LS) for reliability estimation of buried steel water pipeline under the influence of longitudinal and transverse permanent ground deformation. The efficiency of LS has been demonstrated through numerical studies, and the aim of this work is to employ its numerical capabilities in estimating the small failure probability of buried steel water pipeline under the influence of seismic action. To account for the amount and the damage of buried pipe due to seismic action is very difficult due to lack of information and the associated uncertainties. Therefore, LS is employed for the reliability assessment in order to estimate the probability of failure due to seismic actions and associated uncertainty in the structural response. The negative effects of seismic action on buried pipeline are investigated based on the failure modes of excessive tensile and compressive strain.

1. Introduction

In recent time, there is a growing concern about the structural performance of buried steel pipes in areas that are prone to geo-environmental hazards such as earthquake. In designing pipelines for such areas, the main purpose of design engineers is to minimize the probability of failure following an earthquake event. For this reason, there is need to minimize structural damage of buried pipes, in order to sustain the structural integrity of pipelines and prevent emergency failures. In most cases, ground surface structures are more susceptible to the seismic hazards than the underground structures. However buried pipelines at shallow depth may get damaged due to the ground movement.

Uncertainties are involved in the analysis and modelling of buried pipelines in seismic areas. The unreliability may be attributable to the randomness of the physical phenomenon (aleatory uncertainty) and the incomplete knowledge concerning the physics of some of this phenomenon (epistemic uncertainty) (Apostolakis, 1990; Tee and Khan, 2014). Due to these variabilities of the design parameters, the physical or incomplete knowledge of the buried pipelines might develop and the results may lead to failure. However, there is a need to employ a reliable and efficient technique in analysing the probability of failure (Tee et al. 2014; 2015; Khan et al. 2013). For this reason, an advanced simulation method called Line Sampling (LS) is employed for the assessment of failure probability of buried continuous steel pipelines in seismic zones.

This paper presents a reliability assessment of buried pipelines under the influence of longitudinal and transverse permanent ground deformation. Conversely, according to the authors' acquaintance, there is no such work found in the literature on the reliability assessment of buried pipeline under the influence of seismic actions. These effects were analysed for failure possibility regarding excessive tensile and compressive failures.

This paper is organised as follows. Section 2 presents permanent ground deformation (PGD) of a buried continuous pipeline for longitudinal and transverse movement. Section 3 explains the pipe failure criteria and limit states for seismic areas. Section 4 describes the application of LS and the procedure employed for the analysis. Section 5 presents the numerical example. Section 6 discusses the results and significant findings and Section 7 provides the concluding remarks.

2. Permanent ground deformation (PGD)

A substantial number of seismic damages to buried steel pipelines are caused by PGD (Abdoun et al. 2009; Alliance, 2001; Karamanos et al, 2014; O'Rourke and Liu, 2012). PGD can occur as a result of tectonic fault movements, landslides and liquefaction induced lateral spread. The impact of permanent ground deformation depends on the geological formation of the soil. However, designing buried

pipelines for critical responses due to permanent ground deformation (PGD) about continuous pipeline failure is discussed in the following sub-sections.

2.1 Longitudinal permanent ground deformation:

According to O'Rourke and Liu (2012), the pattern of longitudinal permanent ground deformation may be in the form of block pattern, ramp pattern, ridge pattern, ramp-block pattern, and asymmetric ridge pattern. The block pattern of the longitudinal ground movement is uniform throughout the PGD zone and is mostly adapted for a critical response. For a buried pipeline subjected to a block pattern of longitudinal PGD, the maximum axial strain in buried continuous pipelines for both tension and compression can be calculated using Eq. (1) (Dash and Jain, 2007).

$$\varepsilon_a = \frac{T_u L}{2\pi D t E} \left[1 + \frac{n}{1+r} \left(\frac{T_u L}{2\pi D t \sigma_y} \right)^r \right] \quad (1)$$

Where L = the length of PGD zone; T_u = the peak friction force per unit length of pipe at soil pipe interface; D = the outside diameter of pipe; t = pipe thickness; n, r = Ramberg-Osgood parameters; E = the modulus of elasticity of pipe material. The maximum axial soil force per unit length of pipe that can be transmitted to buried pipe can be expressed as shown in Eq. (2).

$$T_u = \pi D \alpha c + \pi D H \bar{\gamma} \frac{1+K_0}{2} \tan \delta \quad (2)$$

Where c = represent soil cohesion of the soil backfill; H = depth to pipe centreline; $\bar{\gamma}$ = effective unit weight of soil; K_0 = coefficient of pressure at rest; δ = interface angle of friction for pipe and soil = $f\phi$; ϕ = internal friction angle of the soil; f = coating dependent factor relating the internal friction angle of the soil to the friction angle at the soil-pipe interface; α = adhesion factor.

$$\alpha = 0.608 - 0.123c - \frac{0.274}{c^2+1} + \frac{0.695}{c^3+1} \quad (3)$$

2.2 Transverse permanent ground deformation

When a buried continuous steel pipeline is subjected to transverse ground deformation, the pipeline will stretched and bend as it attempts to accommodate the applied force. Unlike the longitudinal ground displacement, where the pattern of the ground movement due to transverse PGD can be different. The analytical equations used for this study is based on the expressions given by Dash and Jain (2007), O'Rourke and Liu (1999). The expression is a simplified pipeline response model to spatially distributed PGD. The maximum bending strain in pipe may be conservatively calculated as expressed in Eq. (4).

$$\varepsilon_b = \frac{P_u W^2}{3\pi E t D^2} \quad (4)$$

Where W = width of PGD zone; t = thickness of pipe; P_u = the maximum lateral resistance of soil per unit length of pipe; E = the elastic modulus of pipe material. The maximum lateral soil force per unit length of pipe that can be transmitted to the pipe can be expressed as shown in Eq. (5).

$$P_u = N_{ch} c D + N_{qh} \bar{\gamma} H D \quad (5)$$

Where N_{ch} = horizontal bearing capacity factor for clay (0 for $c = 0$); N_{qh} = horizontal bearing capacity factor (0 for $\phi = 0$).

3 Pipe failure criteria and limit states in seismic areas

3.1 Tension failure

The limit state value of tension strain uses the methods that combine elasticity with plasticity theory. Herein, the reliability analysis of buried steel pipe in seismic areas is based on Alliance's

recommendations (2001). It is suggested that the accumulated strain should not exceed 2.0% in tension.

3.2 Local buckling

Failure due to buckling refers to a state of structural instability where a buried pipe that is loaded in compression experiences an unexpected change from stable to an unstable condition. The pipe can undergo local instability and this could lead to failure due to buckling (wrinkling). After the initiation of local shell wrinkling, further geometric distortion caused by PGD tends to concentrate in areas where there is wrinkle (O'Rourke and Liu, 1999). According to Alliance (2001), the compressive critical strain of local buckling in the longitudinal direction can be estimated using Eq. (6).

$$0.5 \left(\frac{t}{D'} \right) - 0.0025 + 3000 \left(\frac{pD}{2Et} \right)^2 \quad (6)$$

$$D' = \frac{D}{1 - \frac{3}{D}(D - D_{min})} \quad (7)$$

Where D_{min} = the minimum inside diameter of pipe.

4 Application of Line Sampling

Line sampling is a simulation method that is used for computation of small failure probabilities. The main idea of LS is to employ lines instead of random points in order to probe the failure domain of the system that is being analysed (Koutsourelakis et al, 2004; Pradlwarter et al, 2005; Zio and Pedroni, 2010). The procedure employed for estimation of failure probability of a buried continuous pipeline under the influence of seismic actions using LS algorithm can be summarised as follows:

1. The 'important unit vector' that points towards the failure domain is determined using the 'normalised centre of mass' approach (Koutsourelakis et al, 2004; Zio and Pedroni, 2010).
2. Using standard Monte Carlo simulation (MCS), N_T vectors $\{x^k: K = 1, 2, 3, \dots, N_T\}$ with $x^k = \{x_1^k, x_2^k, x_3^k, \dots, x_j^k, \dots, x_n^k\}$ are sampled from the multidimensional joint probability density function $q(\cdot): \mathbb{R}^n \rightarrow [0, 1)$ of the input system.
3. Transform the N_T sample vectors $\{x^k: K = 1, 2, 3, \dots, N_T\}$ defined in the physical space into N_T samples $\{\theta^k: K = 1, 2, 3, \dots, N_T\}$ defined in the standard normal space (Zio, 2013).
4. Estimate one-dimensional failure probabilities $\{\hat{P}^{1D,k}(F): K = 1, 2, 3, \dots, N_T\}$ associated with N_T vectors corresponding to each one of the standard normal samples $\{\theta^k: K = 1, 2, 3, \dots, N_T\}$ obtained in step 3. For each random samples, $\{\theta^k: K = 1, 2, 3, \dots, N_T\}$ the following steps should be performed (Koutsourelakis et al, 2004; Zio, 2013).
 - a. For each sample vectors $\{\theta^k: K = 1, 2, 3, \dots, N_T\}$, determine a vector $\{\tilde{\theta}^k, K = 1, 2, 3, \dots, N_T\}$ as the sum deterministic multiple of the unit important direction and a vector $\theta^{k,\perp}, K = 1, 2, 3, \dots, N_T$ that is perpendicular to the direction of the unit vector for the important direction. This can be expressed as:

$$\tilde{\theta}^k = c^k \alpha + \theta^{k,\perp} \quad (8)$$

Where c^k is a real number with values between $-\infty$ and $+\infty$. The relationship for the perpendicular direction is given as:

$$\theta^{k,\perp} = \theta^k - \langle \alpha, \theta^k \rangle \alpha \quad (9)$$

Where $\theta^k: K = 1, 2, 3, \dots, N_T$ represents a random realisation of the input variables and $\langle \alpha, \theta^k \rangle$ represents the scalar product between the unit vector α and the random variables $\theta^k, K = 1, 2, 3, \dots, N_T$.

- b. Calculate the value of c^{-k} as the intersection between the limit state function $g_{\theta}(\tilde{\theta}^k) = g_{\theta}(c^k \alpha + \theta^{k,\perp}) = 0$ and the line $l^k(c^k, \alpha)$ passing through θ^k and parallel to the important direction α . The value of c^{-k} is determined by fitting a first or second order polynomial and finding its root. For each standard normal random samples, two or three system performance evaluation is required.
- c. Estimate the conditional one dimensional probability associated to each random sample $\theta^k, K = 1, 2, 3, \dots, N_T$. The associated conditional failure probability $\hat{P}_f^k, K = 1, 2, 3, \dots, N_T$ is given as:

$$\hat{P}_f^k = P[N(0,1) > c^{-k}] = 1 - P[N(0,1) > c^{-k}] \quad (10)$$

$$= 1 - \Phi(c^{-k}) = \Phi(-c^{-k}) \quad (11)$$

Where $\Phi(\cdot)$ represents the standard normal cumulative density function.

5. From the result of the conditional one dimensional failure probability $\hat{P}_f^k, K = 1, 2, 3, \dots, N_T$, calculate the unbiased failure probability P_f and its variance as:

$$P_f = \frac{1}{N_T} \sum_{k=1}^{N_T} \hat{P}_f^k \quad (12)$$

$$\sigma^2[P_f] = \frac{1}{N_T(N_T-1)} \sum_{k=1}^{N_T} (\hat{P}_f^k - P_f)^2 \quad (13)$$

5 Numerical example

Consider the determination of reliability of buried continuous pipes under the influence of seismic action leading to longitudinal PGD and transverse PGD, the main purpose of this example is to illustrate the confidence level in analysing the design of buried pipelines in seismic zone. The numerical values of the pipe and soil parameters are given in Tables 1 and 2.

Table 1. Material properties of pipe parameters

Description of parameters	Value	Description of parameters	Value
Pipe diameter D	0.6m	Important factor I_p	1.5
Internal diameter of pipe D_i	0.5872m	Dip angle of fault movement ψ	35°
Expected peak ground acceleration PGA_r	0.45	Angle between pipeline and fault line β	40°
Frictional factor	0.7	Important factor for fault movement I_p	2.3
Ramberg Osgood Parameter n	9	Internal friction angle of soil ϕ	32°
Permanent ground movement parallel to pipe	2	Ramberg Osgood Parameter r	10

Table 2. Statistical random properties of pipe materials

Material properties	Mean	COV (%)	Standard Deviation
Coefficient of cohesion C	30 KPa (Normal)	2.0	0.6 KPa
Soil cover above centre of pipeline H	1.2 m (Normal)	1.0	0.012 m
Unit weight of soil γ_s	18 KN/m ³ (Normal)	2.5	0.45 KN/m ³
Unit weight of pipe material γ_p	78.56 KN/m ³ (Normal)	2.0	1.5712 KN/m ³
Elastic modulus of pipe material E	213.74x10 ⁶ KPa (Normal)	1.0	2.1374x10 ⁶ KPa
Pipe thickness t	0.0064 m (Normal)	1.0	0.000064 m
Yield stress of pipe material σ_y	358 MPa (Normal)	1.0	3.58 MPa

6 Results and discussion

On the basis of constitutive equations from literature (Alliance, 2001; Dash and Jain, 2007; O'Rourke and Liu, 1999), the probability of failure for buried pipeline is assessed based on tensile and local buckling failure using LS and MCS. To evaluate the probability of failure, MATLAB code is developed and implemented. The delivery of the failure probability with respect to tensile strain and

local buckling of buried pipe is calculated by considering longitudinal PGD and transverse PGD. LS has been compared with MC approach using the numerical example in Section 5. A total of $N_T = 450$ samples was used for LS method based on the two conditions of the failure state, whereas a total of $N_T = 100,000$ samples was used for MC approach respectively. The samples used for MC simulation is due to the need of analysing a number of failure samples that is adequate to estimate the associated small failure probabilities with LS. Also the two methods are analysed using unitary coefficient of variation δ_u with an allowable robustness. The sample sizes are considered for the purpose of assessing and validating the probability of failure of buried continuous pipeline due to seismic action. Furthermore it serves as an illustration and comparison between the two methods.

Considering different safe service life of 2%, 2.5% and 3% for tensile failure and 0.00373 for local buckling (compression) failure (Alliance, 2001; Dash and Jain, 2007), the study reveals that seismic action can cause a severe impact on buried pipelines. Tables 1 and 2 report the results obtained for longitudinal PGD and transverse PGD with the procedure above (see Section 4). The outcomes in Tables 3 and 4, also illustrate the efficiency of LS as against MC, since a total number of 450 samples (using LS) is adequate to obtain the probability of failure for 100000 samples (using MC). The small values of unitary coefficient of variation for LS, obtained using the probability of failure demonstrate the advantage of using the abovementioned procedure. It is important to note that if the coefficient of variation is low, the variability of the corresponding failure probability will be small and thus the better the efficiency of the applied method. Based on this concept, it is reported in this study, that LS leads to a substantial improvement in efficiency over MC for the two conditions of failure states due to seismic action.

Finally, the safe service life based on the numerical example for longitudinal PGD and transverse PGD can be evaluated. This study will assist in mitigating against the sudden failure of buried continuous pipes due to the impact of the seismic action. Correspondingly, it is worth noting that the use of favourable lines as a replacement for random points to query the failure domain of interest makes the effectiveness of the LS method almost independent of the target failure probability P_f to be estimated.

Table 3. Values of the failure probability estimate as a result of seismic action of longitudinal PGD

Tension strain limit	MCS		LS	
	P_f	δ_u	P_f	δ_u
0.02	1.0×10^{-4}	46.5	2.2×10^{-3}	2.5
0.025	9.0×10^{-4}	44.3	1.9×10^{-3}	2.2
0.03	8.0×10^{-4}	40.0	1.8×10^{-3}	2.1
Compression strain limit				
0.00373	1.0×10^{-4}	72.8	3.3×10^{-3}	4.1

Table 4. Values of the failure probability estimate as a result of seismic action of transverse PGD

Tension strain limit	MCS		LS	
	P_f	δ_u	P_f	δ_u
0.02	5.0×10^{-4}	79.5	2.6×10^{-3}	0.82
0.025	7.0×10^{-4}	77.6	1.6×10^{-3}	0.94
0.03	1.1×10^{-3}	73.3	2.2×10^{-3}	1.07
Compression strain limit				
0.00373	5.0×10^{-3}	55.3	1.3×10^{-3}	1.60

7 Conclusion

Line Sampling and Monte Carlo simulation approach has been utilised for reliability estimation of buried continuous pipelines subjected to seismic actions due to longitudinal and transverse PGD. The results of LS have been compared with MCS for estimation of small failure probability. MC has been

chosen for validation of the results because of its popularity in assessment of probability of failure. The results show that LS is efficient in estimating small failure probability as against MC. LS resolves the problem of computing the multi-dimensional failure probability integral in the original space by transforming the random variables in the original 'physical' space into a 'standard normal space' where each random variables is represented by an independent central unit Gaussian distribution. The exceptional performance of LS in estimating small failure probability makes it more attractive and a management tool for the assessment of buried pipelines under the influence of seismic actions.

References

- Abdoun TH, Ha D, O'Rourke MJ, Symans MD, O'Rourke TD, Palmer MC, Stewart HE (2009) Factors influencing the behavior of buried pipelines subjected to earthquake faulting. *Soil Dynamics and Earthquake Engineering*, **29**(3), 415-427.
- Alliance AL (2001) *Guidelines for the Design of Buried Steel Pipe*. ASCE, USA.
- Apostolakis G (1990) The concept of probability in safety assessments of technological systems. *Science*, 250(4986), 1359-1364.
- Dash SR, Jain SK (2007) *IITK-GSDMA Guidelines for Seismic Design of Buried Pipelines: Provisions with Commentary and Explanatory Examples*. National Information Center of Earthquake Engineering, Kanpur, India.
- Karamanos SA, Keil B, Card RJ (2014) Seismic design of buried steel water pipelines. *Proceeding of the Conference on ASCE Pipelines*, pp.3-6.
- Khan LR, Tee KF, Alani AM (2013) Reliability-based management of underground pipeline network using genetic algorithm. *Proceeding of the 11th International Probabilistic Workshop*, Brno, Czech Republic, November 6-8, pp.159-170.
- Koutsourelakis PS, Pradlwarter HJ, Schuëller GI (2004). Reliability of structures in high dimensions, Part I: Algorithms and applications. *Probabilistic Engineering Mechanics*, **19**(4), 409-417.
- O'Rourke MJ, Liu X (2012) *Seismic Design of Buried and Offshore Pipelines*. MCEER Monograph MCEER-12-MN04.
- O'Rourke MJ, Liu X (1999) *Response of Buried Pipelines Subject to Earthquake Effects*. MCEER Publications.
- Pradlwarter HJ, Pellissetti MF, Schenk CA, Schuëller GI, Kreis A, Fransen S, Calvi A, Klein, M. (2005) Realistic and efficient reliability estimation for aerospace structures. *Computer Methods in Applied Mechanics and Engineering*, **194**(12-16), 1597-1617.
- Schuëller GI, Pradlwarter HJ, Koutsourelakis PS (2004) A critical appraisal of reliability estimation procedures for high dimensions. *Probabilistic Engineering Mechanics*, **19**(4), 463-474.
- Tee KF, Khan LR (2014) Reliability analysis of underground pipelines with correlation between failure modes and random variables. *Journal of Risk and Reliability*, IMechE, **228**(4), 362-370.
- Tee KF, Khan LR, Li HS (2014) Application of subset simulation in reliability estimation for underground pipelines. *Reliability Engineering and System Safety* **130**, 125-131.
- Tee KF, Khan LR, Coolen-Maturi T (2015) Application of receiver operating characteristic curve for pipeline reliability analysis. *Journal of Risk and Reliability*, IMechE, **229**(3), 181-192.
- Zio E (2013) *The Monte Carlo Simulation Method for System Reliability and Risk Analysis*. Springer.
- Zio E, Pedroni N (2010) An optimized Line Sampling method for the estimation of the failure probability of nuclear passive systems. *Reliability Engineering & System Safety*, **95**(12), 1300-1313.

Carriageway Deterioration Prognosis Modelling using Markovian Chain

Ejiroghene Onome Ekpiwhre^{*}, Kong Fah Tee^{**}, Owen Mordi^{***} and Tony Bull^{****}

^{*}Department of Engineering Science, University of Greenwich, UK: e.ekpiwhre@gre.ac.uk

^{**}Department of Engineering Science, University of Greenwich, UK: k.f.tee@gre.ac.uk

^{***}Department of Highway, Ministry of Works, Asaba, Delta State Government, NG: o.mordi@mow.dts.gov.ng

^{****}Department of Surface Strategy & Planning, Transport for London, UK: tony.bull@tfl.gov.uk

Abstract: Condition deterioration prognostic method for assessing states of highway infrastructures is an essential tool for highway authorities and owners for the maintenance and management of their assets. Carriageway degradation and deterioration from the effect of defects is investigated using Markovian chain buttressed by statistical similarities. This algorithm demonstrated in the report portrays a Linear Transition Probability (LTP) Markovian Chain in an Exploratory Data Analysis (EDA) graphical domain. The application takes advantage of an Automated Condition Survey (ACS) technique. As an illustration, an application with real-time principal classified road (A) inspection data generated from an ACS named SCANNER (Surface Condition Assessment of the National Network of Roads) survey conducted for two consecutive years without maintenance intervention in England is offered. The captured prognostic deterioration condition reflecting a 30-years period developed using the LTP-EDA model is presented therein. The proposed model having a hybrid effect creates better Structural Health Monitoring (SHM) process and good decision making for highway practitioners.

1. Introduction

Highway infrastructure authorities and owners require knowledge of their asset condition for timely and appropriate repairs towards decreasing its adverse effect on road users, reduce further deterioration and optimise whole life cycle cost. In a free seminar championed by the Chartered Institute of Highway and Transportation (CIHT), London, the director, Asset Management (AM), Transport for London (TfL), delivered in a presentation “Asset Management for Londoners by Londoners: Whole life Asset Management of road infrastructures”. Emphasized in the presentation, is the need for the understanding of whole life cycle (WLC) of road assets as it affects areas of safety, function, satisfaction, environment and cost.

The intricate failure mechanism and desperate deterioration trend make modelling deterioration challenging even from historical data (Khan et al 2013). This challenge is addressed herein by the LTM-EDA model developed, having the ability to optimise reliability and predict imminent risk from degrading highway assets. Optimised safety and Risk Based Inspections (RBI) have sustained reliability of civil infrastructures from degradation (Tee et al. 2014, Tee and Khan 2012). This manuscript, Reliability Based Maintenance (RBM) toolkit which addresses inspection maintenance for highway infrastructures, trails a report by Ekpiwhre and Tee (2015) discussing asset functional failures using service inspection datasets. In addition to this opening section, other parts are 2- Background, 3-Deterioration modelling, 4-Stochastic Markov Chain Prognosis, 5-Exploratory Data Analysis and 6- Discussion, Future Work and Conclusions.

2. Background

In the attempt to address the effect of increasing growth in traffic on the road networks in England, the Department for Transport (DfT) published a new Code of Practice (CoP) substituting the 1983 release. This new CoP, DfT (2005), titled ‘well-maintained highways’ acts as a maintenance and inspection mechanism to highway practitioners. The Nigerian Federal Ministry of Work (FMW) laments in print in a Compendium report on road infrastructure FMW (2013), on the consequence of its poorly maintained highway network. The summary estimates the financial impact of its decaying road infrastructure is totaling 175 Billion Naira (N), equivalent to 605,723M in UK Pound Sterling (£) and 870.188M in US Dollars (\$). The losses are N88B ≈ [£304,592M ≈ \$437,580M] from vehicle operational cost, N12B ≈ [£41,535,300M ≈ \$59,670M] from delayed turn around/increased travel time and N75B ≈ [£259,596M ≈ \$372,938] from reduction in asset

value. The metamorphosis between a well or poorly maintained highway as expressed by the CoP and experienced by the Compendium (FMW 2013) is the proper practice of inspection and maintenance implementation. Inspection and survey for maintenance as defined by the CoP categorises inspection for maintenance into three categories, viz. safety inspection, service inspection, and conditional survey.

In the search for better measures for the maintenance condition of roads within England in consideration for more reliable ACS assessment, a new SCANNER specification which replaces the 2006 issue, and its subsequent updates in 2007 and 2009 (DfT 2011). This ACS survey type is likely to displace visual inspection styles in use for roads because of its mechanical advantages. The key SCANNER parameters portrayed as expressed in Table 1 are longitudinal profile, transverse profile, edge condition, texture depth and cracking measured and generated by the moving automated vehicle, JacobsRST26, WDM RAV4 and Yotta (DCL) ARAN1. Although roads have benefited from the introduction of SCANNER surveys for SHM, forecasting deterioration using SCANNER survey data is yet exploited. A challenge faced by highway practitioners having historical data is the inability to conduct an EDA, an approach to data analysis that allows data reveal their underlying structure and models. This methodology, which is similar to statistical graphics, portrays real understanding to SCANNER survey data (Croarkin and Tobias 2006). Forecasting carriageway deterioration using EDA on Markovian model to SCANNER dataset can open new understanding for its SHM seeing the accuracy and high quality of its mobile data set produced.

3. Deterioration modelling

Forecasting degradation of physical assets often uses quantitative models in estimating their decline with regards to the performance (Tee et al. 2015). These quantitative models are often probabilistic or deterministic in nature, having either a discrete or continuous performance measures. The Markov Chain model developed by Andrey Markov, who summarised his findings in 1913, is considered to be a stochastic process because they are discrete in time, having an infinite state space. Some studies recognised Markov Chain as an approach for the simulation of an asset over time using an asset current situation and transition probability (Remenyte-Prescott and Andrews 2013, Welllge 2014).

3.1 The current state (CS)

The current status probabilities of the system $\{S_t\}$ as denoted in Eq. (1) is vital in Markov model simulation.

$$\{S_t\} = \{B_1, B_2, B_3, \dots, B_n\}^t. \quad (1)$$

where $\{S_t\}$ its stationary distribution over time and $\{B_1, B_2, B_3, \dots, B_n\}^t$ represents the determined percentage of various condition band. The variables represent condition bands as follows B1 is Very Good (VG), B2 is Good (G), B3 is Fair (F), B4 is Poor (P), and B5 is Very Poor (VP). The stationary distribution for each band $\{S_t\}$ is obtained via percentage conversion of distribution states by the total calibrated sample.

3.2 The transition probability (TP)

The transition probabilities, $P(ij)$ can be summarised in a matrix of n states, having a probability in each row tallying up to one called the transition matrix. A system when at its worst state VP, is seated at the bottom right end of the Transition Probability Matrix (TPM) $P_{nn}=1$, as articulated in Eq (2). The Highways Maintenance Efficiency Programme (HMEP) encourages the utilisation of LTP because of its decoupling ability showing spread over time (HMEP 2012). The LTP approach, which uses a replicative TP, is assumed suitable for creating a new or initialising inspection and maintenance regimes.

$$P(ij) = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ P_{21} & P_{22} & \dots & P_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ P_{n1} & P_{n2} & \dots & P_{nn} \end{bmatrix} \leftrightarrow \begin{bmatrix} P_{11} & P_{12} & P_{13} & \dots & P_{1n} \\ 0 & P_{22} & P_{23} & \dots & P_{2n} \\ 0 & 0 & P_{33} & \dots & P_{3n} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (2)$$

3.3 The Markovian modelling process

In Markovian process, the CS and the TP express the simulation process as shown in Eq (3)

$$\{S_{t+1}\} = [P] * \{S_t\}. \quad (3)$$

where $\{S_{t+1}\}$ is the one-step TP, $[P]$ is the TPM and $\{S_t\}$ is the CS. The Nth step transition probability vector is a repetition of the process given by Eq. (4).

$$\{S_n\} = [P] * \{S_t\}^n. \quad (4)$$

4 Piloted Markovian chain prognosis using SCANNER survey measurements

The Markovian model described consists of a CS and LTP. The formulation of the CS and LTP originated from SCANNER ACS dataset, and the resultant TP prognosis is discussed below.

4.1 Derivation of current state distribution

The presently developed state utilises defect assessment and road condition indicator threshold as recommended by the DfT (2011) and redefines five-band road condition indicator (RCI) threshold as illustrated in Table 1. The three-band threshold employed by the specification DfT (2011) is perceived not to show clarity in deterioration spread thus the creation of the proposed five-band threshold. The percentage distribution of result generated represents CS, also known as the initial state row vector as shown in Eq. (5) logically.

$$P_{CS} = Q_{CB}/Q_{CS}. \quad (5)$$

where proportion of CS is P_{CS} and Q_{CB}/Q_{CS} is the quantity of condition band and current state, respectively. The populated CS probability distribution vector summing up 1 derived are $P_{CS}(VG)=0.8566$, $P_{CS}(G)=0.1337$, $P_{CS}(F)=0.0097$, $P_{CS}(P)=0$ and $P_{CS}(VP)=0$.

Table 1: Tabulated link of SCANNER and developed RCI threshold.

Carriageways Defect		Threshold For Assessing Condition Bands						
		SCANNER		Applied Five-Band Starting Point				
	Defects Code	Low	High	VG	G	F	P	VP
1	Ride Quality (LV3) -mm ²	< 4	> 10	0 To ≤ 2	> 2 To ≤ 4	> 4 To ≤ 10	>10 To ≤ 12	> 12
2	Ride Quality (LV10)-mm ²	< 21	> 56	0 To ≤ 10.5	>10 To ≤ 21	> 21 To ≤ 56	> 56 To ≤ 66.5	>66.5
3	Rut Depth (LLRD) -mm	< 10	> 20	0 To ≤ 5	> 5 To ≤ 10	> 10 To ≤ 20	> 20 To ≤ 25	> 25
4	Rut Depth (LRRD) -mm	< 10	> 20	0 To ≤ 5	>5 To ≤ 10	> 10 To ≤ 20	> 20 To ≤ 25	> 25
5	Cracking (LTRC) -%	< 0.15	> 2	0 To ≤ 0.075	>0.075 To ≤ 0.15	> 0.15 To ≤ 2	> 2 To ≤ 2.075	>2.075
6	Texture (LLTX) -mm	> 0.6	< 0.3	≥ 0.75	0.75 To ≥ 6	< 0.6 To ≥ 0.3	< 0.3 To ≥ 0.15	≤ 0.15

4.2 Development of linear transition probability

The LTP formulated uses only calibrated data from the datasets sample in developing the change between inspection cycles. The encouraged pair-wise linear deterioration technique by HMEP (2012) aligns herein with the redefined RCI threshold in categorising the proportion of the asset into bands in Table 1. The decline from the Previous State (PS) to the CS ($P_{PS \rightarrow CS}$) within a transition cycle is obtained using Eq. (6) and results generated are $P_{PS \rightarrow CS}(VG)=0.726$, $P_{PS \rightarrow CS}(G)=0.0243$, $P_{PS \rightarrow CS}(F)=0.031$, $P_{PS \rightarrow CS}(P)=0$, $P_{PS \rightarrow CS}(VP)=0$ for the LTP development. Q_{PS} and Q_{CS} are the quantity of PS and CS.

$$P_{PS \rightarrow CS} = Q_{PS \rightarrow CS} / Q_{PS} \cdot \quad (6)$$

Table 2: Linear-rate probability transition from PS to CS.

Bands	VG	G	F	P	VP
VG	0.726	0.243	0.031	0	0
G	0	0.726	0.243	0.031	0
F	0	0	0.726	0.243	0.031
P	0	0	0	0.726	0.274
VP	0	0	0	0	1

The LTP in Table 2 is used to design the LTP %. The Picturesque explains the transition mechanism of remnants or transit quantity within condition bands after a change cycle pictorially. The developed real-time dataset linear transition pattern is from between two preceding states. Obtaining the continuing deterioration uses constant linear iteration change across all five bands as shown in Figure 1. Every transition which contains remnants or transits sums up to 100%.

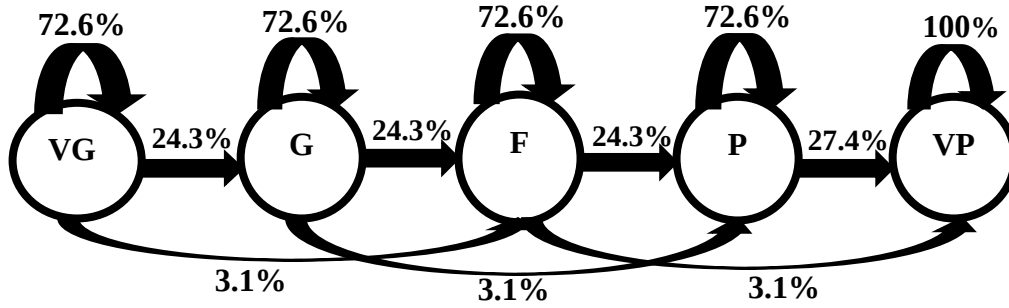


Figure 1. Picturesque linear transition probability of a Markovian Chain Process

4.3 Transition probability prognosis

Future one-step transition process obtainable by multiplying the CS and the LTP as expressed in Eq (7) is used to simulate the FS1 prognosis application in Eq (8) expressing future of 1st-year transition as shown in Eq 9.

$$P_{(FS)} = P_{(CS)} * P_{(LTP)} \cdot \quad (7)$$

$$P_{(FS-1)} = \begin{bmatrix} VG & G & F & P & VP \\ 0.8566 & 0.1337 & 0.0097 & 0 & 0 \end{bmatrix} * \begin{bmatrix} VG & G & F & P & VP \\ 0.726 & 0.234 & 0.031 & 0 & 0 \\ 0 & 0.726 & 0.234 & 0.031 & 0 \\ 0 & 0 & 0.726 & 0.234 & 0 \\ 0 & 0 & 0 & 0.726 & 0.742 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \quad (8)$$

$$P_{(FS-1)} = \begin{bmatrix} VG & G & F & P & VP \\ 0.6219 & 0.3052 & 0.0661 & 0.0065 & 0.0003 \end{bmatrix} \cdot \quad (9)$$

The Markov discrete probability model has the capability to predict more than one performance level using a constant TPM. The Nth transition is simulated using Eq (10), generating linear Nth pair-wise steps of 5, 10, 15, 20, 25 and 30 years is presented in Table 3. These pair-wise steps could be altered as desired as well as age prognosis could be determined until the deterioration reaches a steady state.

$$P_{(FS)} = P_{(CS)} * P_{(LTP)} \cdot \quad (10)$$

Table 3: Tabulated prognoses quantification from CS into Future State (FS) at intervals of 5-years threshold.

N	Future Years	VG	G	F	P	VP
1	CS	0.8566	0.1337	0.0097	0	0
2	FS-1	0.6219	0.3052	0.0661	0.0065	0.0003
3	FS-5	0.1728	0.3161	0.2775	0.1534	0.0806
4	FS-10	0.0348	0.1221	0.2091	0.2327	0.4017
5	FS-15	0.007	0.0364	0.0928	0.155	0.7092
6	FS-20	0.0014	0.0097	0.0329	0.0733	0.8831
7	FS-25	0.0003	0.0024	0.0103	0.0287	0.9583
8	FS-30	0.0001	0.0006	0.003	0.001	0.9864

5. Exploratory data analysis of deteriorated prognosis

LTP simulated results in Table 3 are applied using EDA approach and its graphical spectacle for the deterioration prognosis and Individual-Moving Ratio (I-MR) are presented in addition to its obtained statistical inferences of the status bands expressed in the 30 years deterioration plot in Figure 2(A). This graphical deterioration prognosis shows the connection of the structural health performance decline within and across the different categories. The (I-MR) chart in Figure 2(B) presents the degradation process control defining its mean and the standard deviations. Also, the changes which affect the process are calculated independently for each state from the centre line and the control limits of the I-MR.

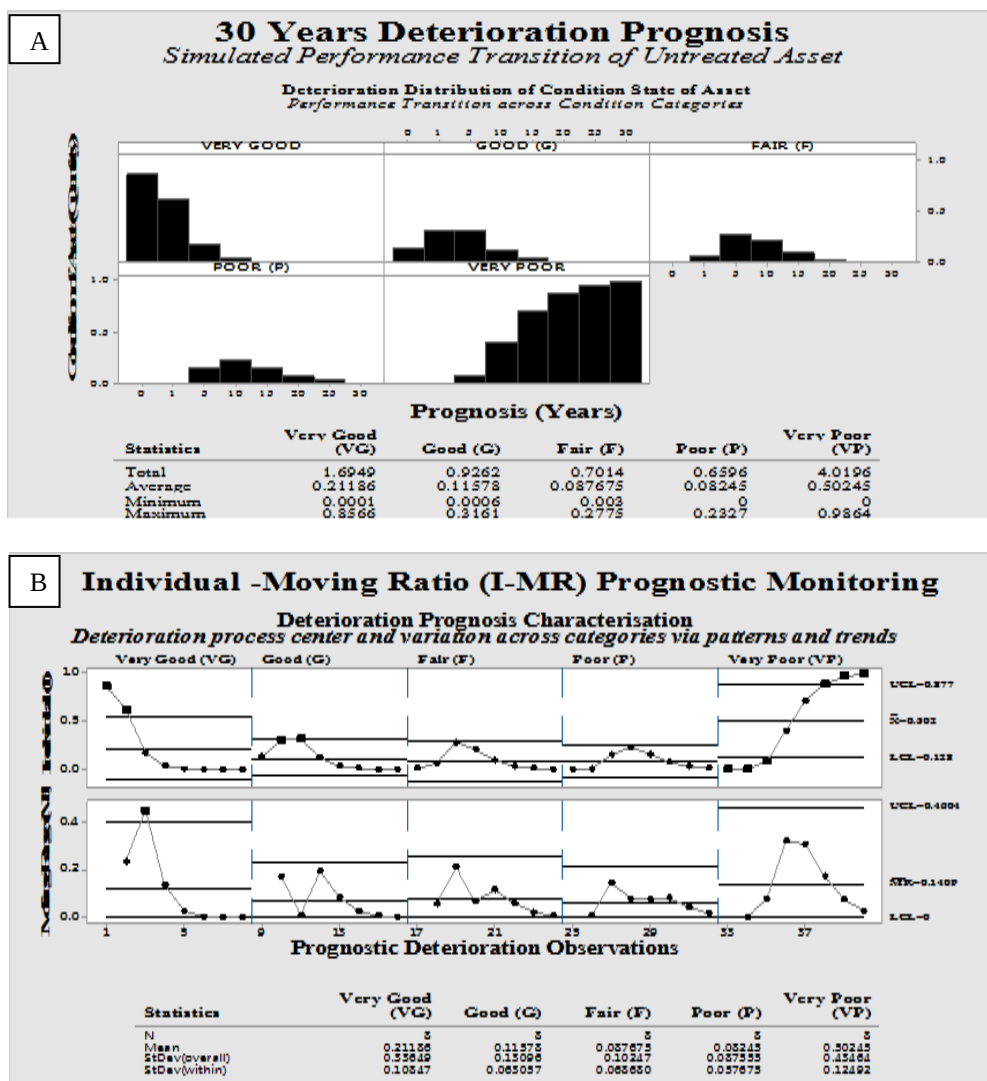


Figure 2. Picturesque EDA's of LTP for five condition bands (A) 30 years deterioration prognosis transitions (B) Prognostic monitoring process in I-MR.

6. Discussion, future work, and conclusion

The LTP-EDA model presents itself suitable for revealing the prognosis of physical assets. Its exploration forecast helps to identify trends, patterns, and comparisons between different deteriorated conditions. The pair-wise decoupling ability makes it possible to show expression in any years pair-wise prognosis period desired. The I-MR observations can measure and monitor traditionally difficult or impossible degradation process. Further work, the ongoing developed RBM toolkit proposes RBI in a stochastic domain with the hope it can augment challenges faced utilising non-motorised inspection datasets. The LTP-EDA model currently restricts on nonclear-cut datasets having inconsistently, thus a caution use for visually inspected road asset. It is also advised to use with caution to roads assets where visual inspection is in practice and systems such as foot/cycleway, structures, drainage systems, traffic systems, signs, street lighting and road making/studs. It can comfortably accommodate other ACS types as it is more responsive to the programmed dataset. Finally, the authors perceive this versatile model can alleviate challenges revealed by the Compendium and enhance the CoP safe practices expressed in Section 2 of this report considering its relaxed set of rules. The rules make it applicable for utilisation to new inspection regimes as well as optimising established inspection, maintenance administrations and WLC of road assets.

Acknowledgement

Authors acknowledge sponsors [Niger Delta Development Commission (NDDC) and the Delta State Government (DTSG)] for the research funding. Thanks are also due to AM director & team of TfL for ACS.

References

- Croarkin C, Tobias, P (2006) e-handbook of statistical methods.
- DfT (2005) Well-maintained Highways: Code of Practice for Highway Maintenance Management. Department of Transport, *The Stationery Office*, London.
- DfT (2011) SCANNER Surveys for local roads: User Guide and Specification. Using SCANNER Survey results, Department of Transport, *Transport Research Laboratory Limited*, Vol.3, London.
- Duling M, Jacobus J (2007) Towards the development of transition probability matrices in the Markovian model for the predicted service life of buildings. Doctoral dissertation, University of Pretoria.
- Ekpiwhre EO, Tee KF (2015) Reliability-Centered Maintenance of Road Junction Transport Assets, *In Proc. of Fifteenth International Conference on Civil, Structural and Environmental Engineering Computing*, (J. Kruis, Tsompanakis and Topping Eds.) Civil-Comp Press: Scotland.
- FWA (2013) Compendium report on road infrastructural and related development in Nigeria- An inventors manual. Federal Ministry of Works, *Pison-REIPPS.DE*
- HMEP (2012) Work Package – Deterioration Model Methodology, Position Statement, *Highway Maintenance Efficiency Programme*, pp.1–5.
- Khan LR, Tee KF, Alani AM (2013) Reliability-Based Management of Underground Pipeline Network Using Genetic Algorithm, *In Proc. 11th International Probabilistic Workshop, Brno*, pp. 159–170.
- Remenye-PreScott R, Andrews J (2013) Review of Infrastructure Asset Management Methods for Networked Systems. *In Proc. 19th Advances in Risk and Reliability Technology Symposium*.
- Skelly D (2015) Asset management for Londoners by Londoners: Making the most of our asset. *In Proc of Free Seminar on Cracking up or Cracking on: Whole life asset management of highways*, Chartered Institution of Highways and Transportation (CIHT), London.
- Tee KF, Khan LR (2012) Risk-Cost Optimisation and Reliability Analysis of Underground Pipelines, *In Proc. Sixth International ASRANet Conference*, London.
- Tee KF, Khan LR, Chen HP, Alani AM (2014) Reliability based life cycle cost optimization for underground pipeline networks, *Tunnelling and Underground Space Technology*, 43, pp. 32–40.
- Tee KF, Khan LR, Coolen-Maturi T (2015) Application of Receiver Operating Characteristic Curve for Pipeline Reliability Analysis, *In Proc. of the Institution of Mechanical Engineers, Part O : Journal of Risk and Reliability*, 229(3), pp. 181–192.
- Wellalage NK, Zhang T, Dwight (2014) Calibrating Markov chain-based deterioration models for predicting future conditions of railway bridge elements. *Journal of Bridge Engineering*, 20(2).

A data quality metrics hierarchy for reliability data

Ralf Gitzel*, Simone Turrin**, Sylvia Maczey***, Shaomin Wu**** and Björn Schmitz*****

*ABB Corporate Research, ralf.gitzel@de.abb.com

** ABB Corporate Research, simone.turrin@de.abb.com

*** ABB Corporate Research, Sylvia Maczey, sylvia.maczey@de.abb.com

**** University of Kent, Shaomin Wu, S.M.Wu@kent.ac.uk

***** Karlsruhe Institute of Technology, Björn Schmitz, bjoern.schmitz@kit.edu

Abstract. In this paper, we describe an approach to understanding data quality issues in field data used for the calculation of reliability metrics such as availability, reliability over time, or MTBF. The focus lies on data from sources such as maintenance management systems or warranty databases which contain information on failure times, failure modes for all units. We propose a hierarchy of data quality metrics which identify and assess key problems in the input data. The metrics are organized in such a way that they guide the data analyst to those problems with the most impact on the calculation and provide a prioritized action plan for the improvement of data quality. The metrics cover issues such as missing, wrong, implausible and inaccurate data. We use examples with real-world data to showcase our software prototype and to illustrate how the metrics have helped with data preparation. This way, analysts can reduce the amount of wrong conclusions drawn from the data due to mistakes in the input values.

1 Introduction

Many critical business decisions in equipment-heavy industries require a good understanding of the reliability of these assets. However, if the reliability is calculated using field data, any data quality problems will affect the quality of the results. In this paper, we briefly describe the key data quality issues in reliability data (Section 2) and show that there is currently no good way to detect them (Section 3). We propose a series of metrics (Section 4) as well as a 3-tiered hierarchy which simplifies the discovery of the most relevant issues (Section 5). We use an example to illustrate our approach and discuss its practical use. The paper concludes with an outlook on future work (Section 6).

2 Reliability data and its data quality issues

In order to estimate the reliability of a product, a series of key data elements is needed. These are often found in one or more tables. However, for the sake of generalization we assume that data consists of multiple properties associated with events and units (e.g. individual motors, machines, etc.). The events in turn are associated with units. The *core failure data* needed are start date, failure date, and an ID characterizing the unit for tying the two dates together. Also, the product type or version needs to be known to properly group the failure data. Additionally, we need *censoring data* for products which have not failed. Sometimes, we can impute missing core data through *substitution data* (e.g. a delivery or production date instead of a start date). Finally, there is a series of data sets which provide *added value* such as failure modes and information that allows a clustering such as application, customer, or even stress levels.

Missing and *recognizably wrong data* are the most obvious data quality issues. Data is recognizably wrong, if it violates some universally accepted rule, natural law (absolute zero temperature) etc. Both missing and recognizably wrong data, most importantly, cause information loss as events or even whole units have to be removed. Wrong data which cannot be detected can distort the results of the reliability calculation. *Implausible data* is data which is not necessarily wrong but there is a strong suspicion that it might be. It is no easy decision what to do with such data but in general a lot of implausible data means that the data quality is poor. If the *data lacks in richness* of information, there will be a lack of detail that makes the results of a reliability analysis less accurate. For example, accurate failure dates are preferable to just knowing the year of failure. Finally, there are several statistical properties that identify a good *sample selection*, avoiding problems such as bias and small sample size. For a detailed discussion see Gitzel et al 2015.

3 Related work

There is rather large body of papers that addresses the topic of data quality. Data quality is a multi-dimensional problem (Kahn et al 2002, Balou and Pazer 1985) and several dimensional frameworks (see Borek et al 2014, pg. 13 for a survey of frameworks) as well as an ISO standard (Benson 2008) have been developed. Most of these frameworks (with the exception of Ban et al 2008) are based on relatively simple metrics.

However, there is little work which targets data quality in the industrial context, especially with a focus on reliability data analysis. Since data quality is a highly domain-specific topic, this means that there is a large gap in the prior art which needs to be filled. While there are papers which identify data quality problems in industrial data (e.g. IEEE 2007, Vadlamani 2007, Bendell 1988, Montgomery and Hodkiewicz 2014) and there is a supposition that the results of an analysis performed on poor reliability data is “severely misleading” (Bendell 1988), there is little work on quantifying these problems. A notable exception besides our own prior work (Gitzel 2014, Gitzel et al 2015) is a paper on assessing reliability data quality by Montgomery and Hodkiewicz which focuses on the “availability of specific data fields or individual records, the relevance of the data in specific fields, and the accuracy of the data” (Montgomery and Hodkiewicz 2014). Also of note should be algorithms which are designed to deal with “coarse” data, i.e. data of poor quality (for a review, see Wu 2013). A common approach is to use estimations to correct errors (e.g. Ala, and Suzuki 2009). These algorithms can be used in conjunction with our proposed metrics to address the problems discovered.

4 Data quality metrics

In order to understand the data quality problems of a particular case, we use a series of metrics, which represent an updated version of the metrics developed in Gitzel et al 2015. All metrics range from 0 to 100% where a high value means a good quality. Note that some of the names do not follow the literature at the request of our stakeholders.

4.1 Completeness metrics

A common problem is *incomplete data*. Missing data can either be empty properties such as “N/A” or “unknown”. For our data quality analysis we have created a list of such terms. In a database, incomplete data takes the form of empty cells. For each property of each event and unit (see section 2), we need a completeness metric. The metric reflects the percentage of properties which are not empty.

4.2 Free-of-error

Errors in the data are one of the most common data quality problems. While not all errors can be discovered, there are some rules which can be used to discover them. For the purpose of our discussion, we distinguish between *logical*, *set-membership* and *syntactical errors*. For each rule there is a metric which covers one or more properties. The metric is the percentage of units or events that violate a particular rule.

Logical errors can be discovered through rules based on one or more data elements, many of which depend on the domain. Set membership errors occur if a data element should be one of several possibilities but the actual value is not one of those. Syntactical errors are typos that let a value deviate from a given pattern. For example, an event type could be failure or censoring but not “break-down” (set-membership error). As another example, the start date of a product cannot occur after the failure date (logical error). Providing the string “Q2.02.2015” as a date is a syntax error.

4.3 Inconsistency

Inconsistency is a data quality problem of lesser importance. It mostly refers to consistency of notation (e.g. format of serial numbers) or units (cm vs. inches). Inconsistency is not a problem if it does not hide duplicates (e.g. the same unit registered under different serial numbers) or leads to mixing different units of measurement. However, it does increase the probability of such errors occurring. Metrics measuring inconsistency can count the number of elements which conform to a specific pattern, e.g. a predefined structure for serial numbers.

4.4 Plausibility

Like free-of-error metrics, plausibility metrics are based on a series of rules. Many of these rules are highly domain-specific but there are some recurring plausibility issues. For example, there are some basic sanity checks associated with events. Normally, we know the release date of a product, so we can check whether there are any start dates before the corresponding release date, which is not impossible (e.g. pilot installations) but improbable in larger numbers. In the context of events, a common problem is the “*double tap*”, where a failure occurs very shortly (ca. 24h) after another one. In many cases, this is a reporting artifact and needs to be removed.

Finally, there should be one or more rules to check the *distribution of dates*. If there are dates which occur with great frequency, this could be an assumed date for old data. Our metric for this rule is 1 minus the highest percentage of occurrences. Also, if there are large areas devoid of failures, it could be that no records exist for that period of time. If there is a suspicion about such an area, a simple metric can measure the percentage of units which start and fail before or after the gap and are thus not affected. More advanced metrics could discover such voids.

Plausibility can also be applied to *maximum and minimum values* for a particular property. It is possible to have one free-of-error metric which converts impossible values (running hours that exceed the time between installation and failure) and one plausibility metric to identify unlikely values (running hours above x% of the time span as appropriate for the product).

A special kind of plausibility rule covers *sub-population membership*. If you want to base your subpopulation on certain properties (e.g. a specific customer and failure mode) there is a problem if these properties are affected by poor data quality. We use a metric which is 1 minus the percentage of units that might or might not be part of the fleet. This includes empty values as well as names that are similar (as identified by means of a Damerau-Levenshtein algorithm). Often the latter problem is quick to fix once identified.

Data that is *not trustworthy* comes from dubious sources, e.g. data collected by people we do not know for purposes other than reliability analysis. Unfortunately, the metadata required to judge the trustworthiness of data is rarely available. If such data is available, metrics can count all units and events coming from trustworthy sources.

4.5 Richness of information

If there are no details in the information, we consider the information to *lack in richness*. For example, a temperature could be described as “hot” or as “38°C”. Both statements can describe the same temperature but the second description is far more accurate. A typical issue in reliability data is inaccurate dates (production month instead of exact day). In a few cases, we do not have information about individual failures, i.e. we know the number of units that failed in a time period but not which units. In many cases, richness is a static assessment. For example, if there is only a property “*manufacturing year*” instead of “*manufacturing date*”, the richness is low throughout the sample because the exact day within the year is not known. However, it is also possible to search for properties that are suspicious (e.g. “1.1.20xx” or “6’000”). If the majority of running hours entries is multiples of 10, 100, or 1000, the information is probably less rich than numbers which are non-zero in multiple digits. In this case, a metric would measure the percentage of units that are not as accurate as the units of a property would suggest.

Information richness also covers machine readability. For example free-text failure descriptions are less accurately interpretable to a software-based algorithm than failure codes. An issue related to richness is the *excessive use of a category “other”*. While using “other” to cover exotic cases is perfectly acceptable, a property that is “other” in 90% of the cases is quite useless for the purpose of identifying sub-populations.

4.6 Sample selection

In statistical analysis, sample selection is an important quality factor. It is not possible to measure all aspects of sample selection, however, there are a few factors which can be quantified easily. Typically, the **sample size** is of great importance for the quality of the calculated reliability. What makes a good sample size depends on the standard deviation of the population the sample is taken from (which can be estimated from

the standard deviation of the sample). We propose to use a metric which is reduced by the relative standard error of the mean for sample sizes of 30 or more and 0% for sample sizes below 30.

Furthermore, the selection of the sample must be *unbiased*, i.e., the sample should represent the population from which the sample is collected from. In other words, the composition of the sample regarding environmental factors, maintenance policy, applications, industry etc. should be very close or identical to that of the total population. For example, if a sample contains 30% assets from the chemical industry and 70% from food & beverage while the installed base for a product is 50/50, the sample will be unbiased. The metric to measure unbiasedness can be based on the Chebychev distance between the percentages in the sample and the percentages in the population, calibrated to give 100% if the distance is 0.

A special form of bias is *warranty bias*, i.e. samples where the majority of products come from the warranty period. A strong warranty bias implies that the reliability analysis focuses on those products with the poorest performance. A simple metric just counts the percentage of data originating from the warranty period.

Another problem occurs if we only know about failures and not about non-failures. In order to get an unbiased sample, we also have to include non-failures as censored data. However, in some cases (e.g. if we only know warranty data) the censoring date has to be based on an assumption for many or all of the units. We call this phenomenon *unconfirmed censoring* and have found that it can greatly impact on warranty data analysis and needs to be addressed with additional data collection (e.g., questionnaire surveys).

4.7 Substitution quality

Ideally, our reliability calculation uses start dates and failure dates. However, in some cases, this information is not available and we substitute dates which are correlated to these, e.g. delivery dates. If there is a high *level of substitution*, this affects the quality of the calculation. We use a metric that tracks the percentage of units which use the real start date instead of substitutes like the manufacturing date. A second metric tracks the same for failure dates as opposed to reporting dates.

5 Data quality metrics hierarchy

When using our first version of the metrics (Gitzel et al 2015), we found that it was difficult to identify the key problems if there was a lot of less important information in the database. For example, an implausible company name is not as problematic as a missing failure date. Using a hierarchy of metrics allows data analysts to quickly understand the key issues in order to react to them.

5.1 Importance of metrics

Our hierarchy of metrics is based on different levels of importance for data quality problems. We use 5 different levels of importance, as described in the following.

“*Critical*” metrics address problems which will remove an event or unit from the calculation, thus reducing the sample size, or which might falsify the results (an implausible failure date). If any of the critical metrics is low, action is required before the results of the reliability calculation can be used.

“*Substitution for Critical*” metrics identify the degree to which the real critical values are replaced by inaccurate values that need to be used because the real value is not known (see Section 0). These problems are often not worth the effort to fix but improvements in the data collection process can improve the accuracy of the reliability calculation.

“*Subfleet*” metrics cover problems which only become relevant if we want to examine subpopulations. For example, without good data about failure modes, we can hardly identify the reliability curve for a particular one such as “explosion”. Subfleet metrics should be treated as critical metrics once we actually focus on a particular sub-population but only require action if such calculations are desired.

“*Added Value*” metrics cover information that was included but is not directly used in neither the calculation nor the creation of subpopulations. However, such information might be useful to a human scanning the data for some reason, e.g. to identify a unit for further investigation.

The category of “*Other/Unspecified*” metrics only exists for the purpose of “debugging” as it contains any metric that was not assigned to a specific category. In a correctly designed metrics hierarchy, this category is empty.

5.2 Hierarchical approach based on box-plots

In order to quickly track down key problems, we suggest a hierarchy of criticality (see Section 5.1), category (see Section 4), and individual metrics. We represent each level as a boxplot of all metrics in each category.

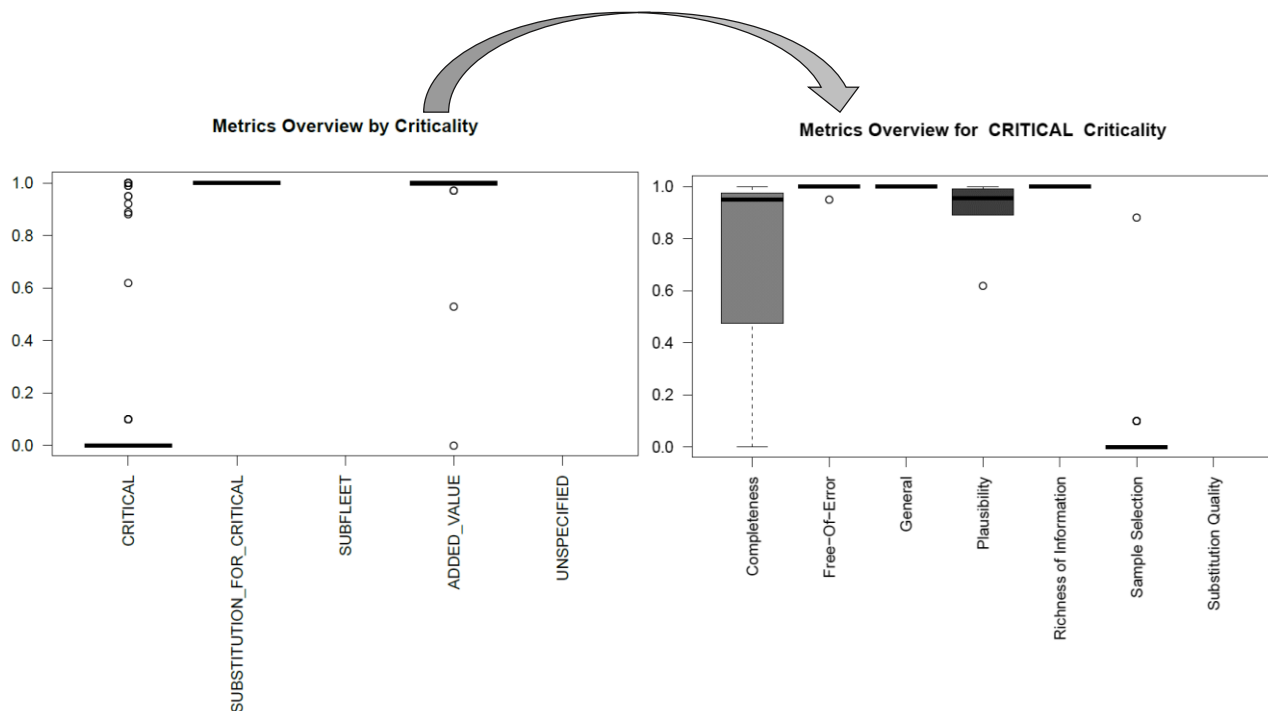


Figure 1: Level 1 and Level 2 Boxplots

Consider the anonymized example shown in Figure 1. The chart on the left shows all metrics, grouped by criticality. (In the example, there are no subfleet and unspecified metrics, so these columns are empty.) We can see at a glance, that the average critical metric is quite poor (as indicated by the bar) with a few metrics being quite good (the outlier dots). If we drill down into a view of all CRITICAL metrics (shown on the right), we find that the key problems are associated with the sample selection and the completeness, with most other metrics being quite good. Sample selection is quite poor on average, completeness metrics on the other hand cover a wide range as indicated by the box plot and its rather long whisker.

A closer investigation of all critical sample selection metrics brings us to the level of individual metrics for a particular category and criticality (see Figure 2). As we can quickly discern, the main problem in this scenario is that for each of the product types the sample size is too small (left). Completeness of data is quite good except for the missing disposal date which is missing for all cases (right).

5.3 Practicality of use

As can be seen by the example, the hierarchy of data quality metrics can be used to quickly identify the key problems in a data set. However, each data set has its unique problems. We have therefore developed a software framework, which can be used to calculate the metrics for a particular data set. Including the configuration of the metrics, it took us two person days to implement a set of metrics tailored to the particular problem. (This does not include the initial assembly of the data.) Using several iterations, we were able to fix multiple problems in the data up to the quality level shown in the example graphics. For example, in the first go, many of the product types were missing, leading to a poor performance of the critical completeness metric.

6 Conclusions and future work

In this paper, we have presented a hierarchy of metrics which allows the quick identification to data quality problems in a data set used for survival analysis. The metrics and their classification are based on several test data sets we had available in our company. The metrics have been implemented in a prototypical API based on Java and R which can be used to quickly program report generators for a particular data set.

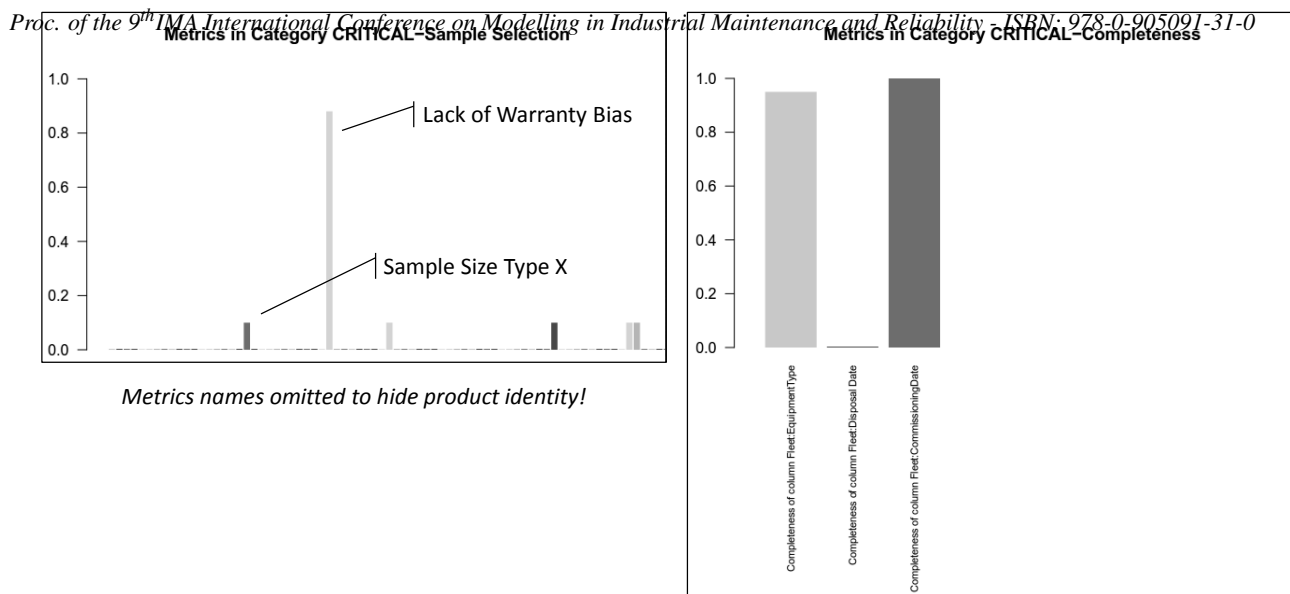


Figure 2: Level 3 Showing Individual Metrics

We plan to further refine our metrics by gaining a better understanding of the impact the data quality has on the results. Currently, the metrics measure the extent of the problem within the data (e.g. how many properties are empty) but not the impact (by how much is the reliability curve distorted by this). A low hanging fruit in this context is to weave the estimated power of the test into the sample selection metrics.

7 References

- Gitzel, R. (2014) *Industrial Services Analytics*. Presentation at the 1. GOR Analytics Tagung, Munich, 11.4.2014
- Gitzel, R., Turring, S., Maczey, S. (2015, July). A Data Quality Dashboard for Reliability Data. In *2015 IEEE 17th Conference on Business Informatics (CBI)*, Vol. 1, pp. 90-97.
- Kahn, BK.; Strong, Diane M.; Wang, Richard Y. (2002): Information quality benchmarks: product and service performance. *Communications of the ACM* 45 (4), pp. 184–192
- Ballou, D.P., Pazer, H.L. (1985): Modeling data and process quality in multi-input, multi-output information systems. *Management Science* 31, 2, (1985), 150–162.
- Borek, A.; Parlikad, AK.; Webb, J.; Woodall, P. (2014): *Total information risk management – maximizing the value of data and information assets*. Morgan Kaufmann
- Benson, P. (2008): ISO 8000 the International Standard for Data Quality, in *MIT Information Quality Symposium*, July 16-17, 2008
- Ban, X., Ning, S., Xu, X.; Cheng, P. (2008): Novel method for the evaluation of data quality based on fuzzy control. *Journal of Systems Engineering and Electronics*, 19 (3), 606–610.
- IEEE (2007): IEEE Standard 493 - IEEE Recommended Practice for the Design of Reliable Industrial and Commercial Power Systems.
- Vadlamani, R. (2007): Modified Great Deluge Algorithm versus Other Metaheuristics in Reliability Optimization, *Computational Intelligence in Reliability Engineering, Studies in Computational Intelligence* 40, 2007, 21-36
- Bendell, T. (1988): An overview of collection, analysis, and application of reliability data in the process industries. *IEEE Transactions on Reliability* 37(2), 132–137.
- Montgomery, N.; Hodkiewicz, M. (2014): Data Fitness for Purpose. In: *Proceedings of the MIMAR 2014 Conference*, Oxford, UK.
- Wu, S. (2013): A review on coarse warranty data and analysis, *Reliability Engineering & System Safety*, 114, 1-11
- Alam, M.M.; Suzuki, K. (2009): Lifetime Estimation Using Only Failure Information From Warranty Database, *IEEE Transactions on Reliability*, 58(4), 573–582

A two-phase life prediction model of explosion proof valve of lithium ion battery under continuous variable-stress

Wenjun Gong*, Yunxia Chen**, Wei Zhang*** and Rui Kang****

* School of Reliability and Systems Engineering, Beihang University, Beijing, China, gongwenjun@buaa.edu.cn

** School of Reliability and Systems Engineering, Beihang University, Beijing, China, chenyunxia@buaa.edu.cn

*** School of Reliability and Systems Engineering, Beihang University, Beijing, China, zhangwei.dse@buaa.edu.cn

**** School of Reliability and Systems Engineering, Beihang University, Beijing, China, kangrui@buaa.edu.cn

Abstract. As the critical component of NCM lithium ion battery to guarantee its safety and reliability, Explosion proof valve's life prediction is essential during the design phase of battery. Traditional methods with accelerated life test data can only record failure time, regardless of describing actual damage mechanism under continuous variable-stress. Thus, a new model in consideration of actual deformation process is proposed for life prediction of the explosion proof valve. It can be divided into two phases: the zero damage and accumulation damage phases. Firstly, the initial gas pressure just make explosion proof valve stay in the elastic zone. Once the gas stress surplus the elastic limit, it begins to cause structure into plastic deformation phase. Then, an accumulation damage model should be established. The whole life prediction result of explosion proof valve is generated by adding two parts. One is the time from zero to the critical gas stress that makes structure stress reach at the elastic limit. It can be achieved by the gas release function with stress simulation result of explosion proof valve. The other is the time from the beginning plastic deformation to the damage limit. It can be determined by actual test data. Finally, an example is presented to validate our method and show its advantage.

1. Introduction

NCM (nickel-cobalt-manganese) polymer lithium ion battery is one of the most widely used commercial lithium ion batteries where NCM polymer is utilized as the new generation cathode material (Kim et al., 2008). With many advantages such as small size, high energy density, relative low cost environment friendly (Shaju and Bruce, 2006; Belharouak et al., 2003), it can be applied to electronic devices like computers or smartphones, new energy vehicles, and some other energy storage fields (Shuang et al., 2014). With the growing applications of NCM-based lithium ion battery in public transportation field, safety and reliability are gradually becoming the key factor which restrict its widespread development. Explosion proof valve, as the critical component to guarantee the safety of batteries, can release the inner gas pressure timely that continuously increase caused by the chemical behavior in batteries. The improper life prediction of explosion proof valve can result in catastrophic consequences, like obscure explosions of mobile phones or fire breaking out of new energy vehicles (Balakrishnan et al., 2006). These accidents bring a heated dispute about their utilization in the mass transit systems. It is evident that traditional life prediction methods and general life test of explosion proof valve is very difficult to satisfy the requirements of safety and reliability. Therefore, an effective and accurate test method of explosion proof valve life prediction is desperately needed.

The current methods of explosion proof valve life test are the accelerated life test (ALT). It can only reflect the external environment factors regardless of the specific failure process, without providing the exact information about weak location (Doksum and Høyland, 1992). In addition, the change of the structure performance parameters is not considered by ALT. these deficiencies mentioned above force us to reconsider another method, such as accelerated degradation tests (ADT) or other failure mechanism tests (Peng and Tseng, 2010). Due to better estimation of reliability and life for high-reliable products, ADT method was employed in this paper as a strong supplement to ALT. the explosion strength deteriorating data was utilized to extrapolate the explosion safety valve life at normal working stress. Explosion proof valve of lithium battery is made of aluminum alloy with high elongation, which is widely used in new energy vehicle industries. The particular characteristic of explosion safety valve made of such type aluminum alloy material need to be clearly pointed out is that chemical behavior in the battery leads to the gas pressure increasing over time (Sun et al., 2016; Séronie-Vivien et al., 2012). Once over the deformation threshold, the explosion

proof valve burst to release the gas inside the battery for the sake of safety. Thus, structure works in a continuous deformation span including elastic zone and plastic zone with continuous variable-stress. In order to reduce the risk of life prediction, it is essential to divide the total life period into two phases. Detailed description can be presented in section 3. The rest of this paper is organized as follows. Firstly, based on deformation analysis of explosion proof valve during the gas inflation test, the failure mechanism can be determined that includes elastic and plastic deformation in section 2. Then, a life prediction model is proposed in section 3. Finally, an example is presented to validate our method and show its advantage.

2. Experimental analysis

In this section, the purpose is to demonstrate the actual deformation of explosion proof valve is through two deformation phases. The material of explosion proof valve is a new type of novel aluminum alloy. From the result of standard tensile sample test, we can obtain tensile strength and yield strength of the structure material which are 155MPa and 85MPa respectively. The gas pressure limit of explosion proof valve in design period is 0.6MPa. For demonstrating the actual deformation pattern of structure, a gas inflation test with four samples at 25 centigrade has been conducted (see Fig. 1). The whole procedure of gas inflation can be described as follows. Firstly aerate these four empty batteries slowly, and separately hold on for 30 seconds at each pressure step that is 0.05MPa, 0.1MPa, 0.15MPa..., 0.5MPa. The result shows that detailed deformation process undergoes elastic and plastic deformation phases. It can be described as Fig. 2 shown.

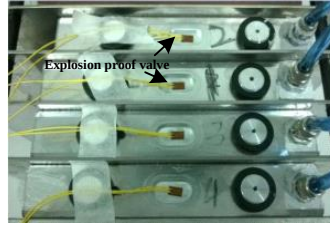


Figure 1 Experimental setting of explosion safety valve

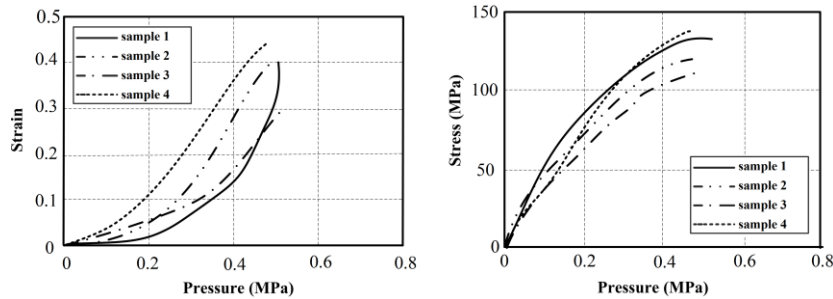


Figure 2 the strain and stress relation curves with gas pressure

3. Life prediction modelling for explosion proof valve

Based on the actual observation, lithium ion battery works under continuous variable-stress. As the typical stress types, temperature and gas pressure exert prominent impacts on the working life of explosion proof valve. Thus, it is crucial to acquire their stress spectrums for accurate life prediction. However, the gas production behavior of battery has a typical time cumulative effect, which causes inner gas pressure presenting a non-linear characteristic. Due to fuzzy uncertainty of gas production at the early stage of gas production, the establishment of corresponding theoretical model is unrealistic. In order to reduce the risk of life prediction, it is essential to divide the total life period into two phases. The first phase is the cumulative part from zero to the initial point P_0 that can cause the actual damage of material, which we call it the elastic deformation phase. It assumes to be lasting for the duration time t_0 . The value of t_0 can be achieved by the fitting gas release function and the simulation stress result of explosion proof valve. The second phase is called the plastic deformation phase; the deformation limitation assumes to be P_{th} . The duration time t_1 can be determined by extrapolation based on ADT data. Therefore, the total life prediction value equals $t_0 + t_1$. Finally, Entire life prediction procedure can be presented in Fig. 3. The next work in this section will focus on the exact calculation process of obtaining the value of t_0 and t_1 .

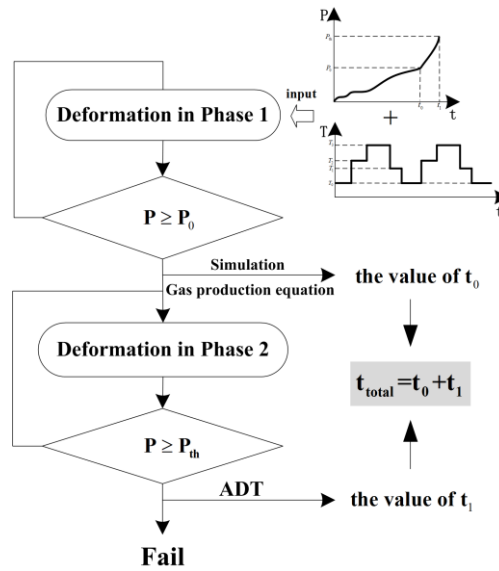


Figure 3 Life prediction procedure of explosion proof valve

3.1 Calculation equation of t_0 and P_0

The gas production process is complexed with numerous interacting reactions, including impacts of external circumstance and inner chemical substance variation (Kumai et al., 1999). Thus, Theoretical method of establishing specific gas production equation is impractical. Ambient temperature is considered as one of the main factor of affecting gas production volume. Thus, we consider detecting the exact gas pressure at different temperature levels by the sensors. The exact variation of gas pressure over time can be depicted as the inner gas production curves. Based on the data of gas production test, a fitting model can be built for the life prediction in the first phase. The generalized Arrhenius model is considered as the basic model which is commonly utilized in occasions covered with gas-phase reactions. It can be obtained as

$$\frac{\partial P}{\partial t} = R(t) = A_t \exp[-(E_a / kT)] \quad (1)$$

Where P represents the degradation character of system, $\partial P / \partial t$ is described as reaction rate, A_t is a constant, k is the Boltzmann constant that equals $8.617 \times 10^{-5} \text{ eV/K}$, E_a represents activation energy.

For the purpose of obtaining parameters value of Eq. (1), inner gas production test data at temperature levels can be utilized. The inner gas pressure variation with time t assumes to be

$$P(t) = P(0) + \frac{\partial P}{\partial t} \cdot t \quad (2)$$

Based on the exact fitting equations at different temperature points, a series of $\partial P / \partial t$ measurement can be obtained. By taking the logarithm form of Eq. (1) that is $\ln(\partial P / \partial t) = a + b/T$ (a, b are parameters to be estimated), the inner gas production model can be determined. Then, given the value of P_0 , the value of t_0 can be determined by Eq.(2).

However, the accurate method of determining the value of P_0 might be unavailable. Thus, in this paper, considering the result of deformation simulation of explosion proof valve, a possible value of P_0 with maximum probability can be obtained. All these processes are presented in section 4.

3.2 Degradation modelling to obtain the value of t_1

First of all, the calculation equation of t_1 is based on the degradation process of endurance strength of metallic materials. It can be obtained as

$$\dot{l} = a \times P^b \times e^{-\frac{c}{T}} \quad (3)$$

Where $\dot{l}$ represents elongation of material, $P(P \geq P_0)$ is the inner gas pressure, $T(T \geq T_0)$ is the surrounding temperature of vent, and a, b, c is the parameter for evaluation.

For evaluating the plastic deformation duration time t_1 , it is necessary to get the boundary of elongation as the failure threshold. However, it is unrealistic to measure continuous increasing process of elongation online at this moment. Based on the principle of Eq. (1) above, the corresponding degradation model can be established. It is assumed that actual stress intensity at time t and temperature $T(t)$ satisfies

$$S(t) = S(t_0) - \int_{t_0}^t A \times (P(t))^B \times e^{\frac{C}{T(t)}} dt \quad (4)$$

Where $S(t)$ represents the bursting strength at time t with the fixed temperature level T , $S(t_0)$ is the initial bursting strength at initial time t_0 . $P(t)$ is the gas pressure at time t . A, B, C are parameters to be estimated.

4. Statistic analysis of test data

According to the test plan mentioned above, the statistic analysis of test data can be divided into three steps. Description of each step is presented in the following part.

4.1 process to determine the value of t_0

Step 1: gas production model based on gas production test data

Firstly, inner gas production test data at temperature levels of 25°C, 60°C, and 70°C is utilized to acquire parameters of the generalized Arrhenius model. Through linear fitting process at these three temperature points, we can get their specific equations as Table 2 described. Comparison with actual inner gas production curves at different temperature levels can be depicted as Fig. 4 shown.

Table 2 the linear fitting equations at different temperature points

Temperature points	Fitting equations	Correlation coefficients
25°C	$P(t) = 0.0003675t - 0.0011346$	0.9869
60°C	$P(t) = 0.00010331t + 0.0074462$	0.9826
70°C	$P(t) = 0.002213t + 0.018304$	0.9608

From fitting equations in Table 2, Different values of $\partial P / \partial t$ at corresponding temperature points can be obtained. Meanwhile, value of a, b can also be determined with the method of least square fitting, that is $\hat{a} = -3748.8, \hat{b} = 11.531$. Corresponding coefficient equals 0.94, which demonstrates that we achieve a reasonable accuracy. Besides, considering that $P(0)$ is a fixed value regardless of temperature's impact in this case, it equals $(-1.1346 + 7.4462 + 18.304) \times 10^{-3} / 3 = 8.2052 \times 10^{-3}$ MPa.

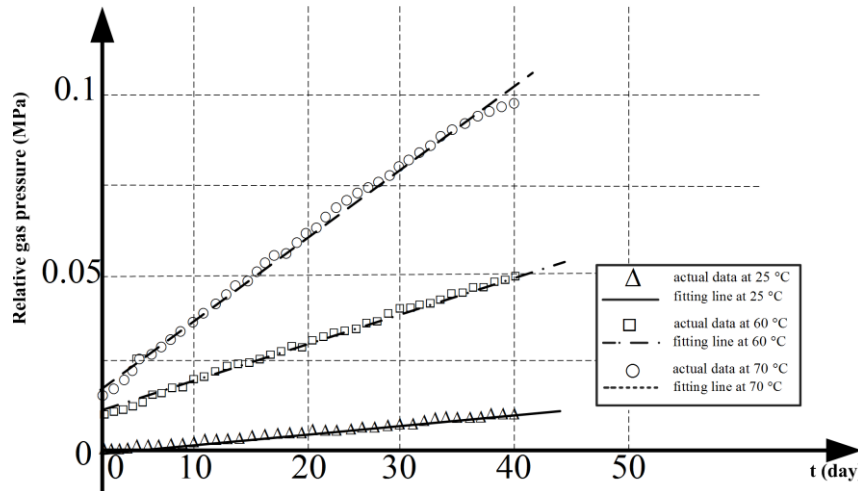


Figure 4 the inner gas production curves

Step 2: determine the value of P_0

In this step, a stress simulation model of explosion safety valve is established to determine the turn point that is the value of P_0 . Simulation result can be presented in Fig. 5, which reflects the actual possible breakdown point during the period of increasing gas pressure. Then, stress and strain simulation curves of the explosion safety valve can be obtained as Fig.6 shown. Based on the assumption of P_0 being the initial point that can cause the actual damage of material, it can be thought that P_0 approximately equals the internal gas pressure corresponding to the elastic limitation 85MPa. Therefore, the estimation value of P_0 is 0.2MPa.

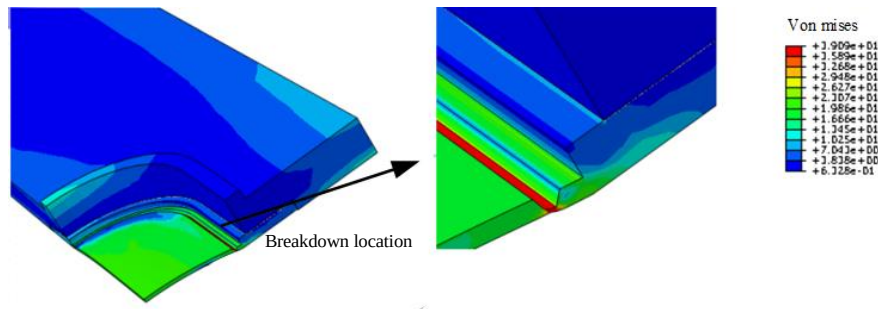


Figure 5 stress simulation result of explosion safety valve with gas pressure increasing

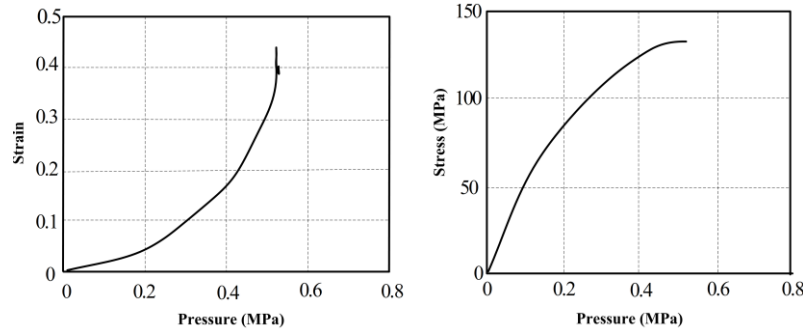


Figure 6 stress and strain simulation curves of the explosion safety valve

Step 3: determine the value of t_0

Combing the result of step 1 and step 2, the value of t_0 at each temperature point can be obtained. Through analyzing surrounding temperature variation on the operation stage, proportion of time occupied at different temperature values in one period can be described as Fig. 7 shown.

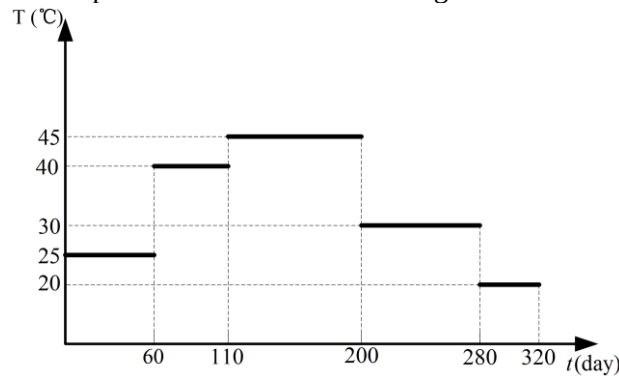


Figure 7 proportion of time occupied at different temperature values in one period

In a period of 320 days, the cumulative gas production pressure P_{cum} can be calculated as

$$\begin{aligned}
 P_{cum} &= 8.2052 + \exp\left\{11.531 - \frac{3748.8}{298}\right\} \times 60 + \exp\left\{11.531 - \frac{3748.8}{313}\right\} \times 50 + \\
 &\quad \exp\left\{11.531 - \frac{3748.8}{318}\right\} \times 90 + \exp\left\{11.531 - \frac{3748.8}{303}\right\} \times 80 + \exp\left\{11.531 - \frac{3748.8}{293}\right\} \times 40 \\
 &= 176.5893 \text{ KPa} = 0.1765893 \text{ MPa}
 \end{aligned}$$

Therefore, based on the estimation value of P_0 that equals 0.2MPa, the value of t_0 can be obtained by $t_0 = 320 + 43.4027 = 363.4027$ (days).

4.2 process to determine the value of t_1

In this part, a degradation test has been considered, the actual degradation trends of explosion strength can be described as Fig. 8 shown.

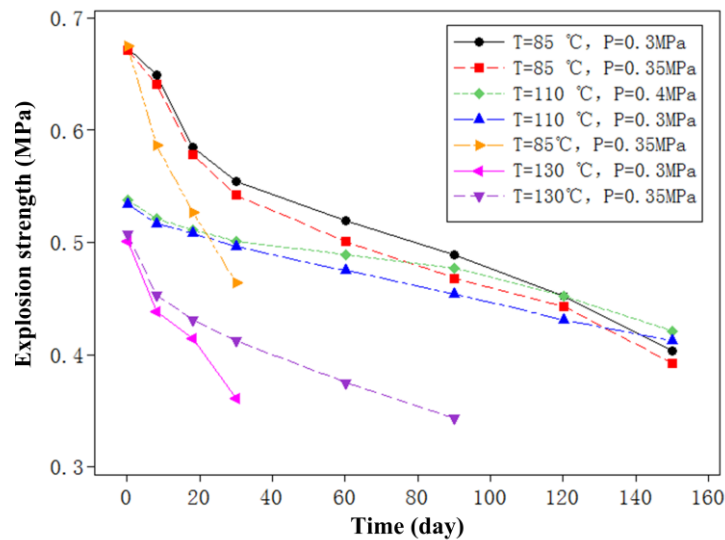


Figure 8 Degradation processes of explosion strength at different environment stress

It is obvious that initial explosion strength is relevant to temperature levels. Thus, a polynomial fitting model of initial explosion strength can be obtained as $P_T(0) = -0.00002T^2 + 0.0105T - 0.7267$. In addition, a mean estimation values of parameters A, B, C of Eq. (4) can be determined, that is $\hat{A} = 8.16898 \times 10^{-8}$, $\hat{B} = 4.526384596$, $\hat{C} = 5709.669913$. Then, a degradation process of the explosion proof valve and gas production curves during the whole life period can be obtained as Fig. 10 shown. And we can calculate the value of t_1 equals 645.8932 days.

In final, the total life prediction value of explosion proof valve t_{total} equals the sum value of t_0 and t_1 that is $t_{total} = t_0 + t_1 = 363.4027 + 645.8932 = 1009.2959$ (days).

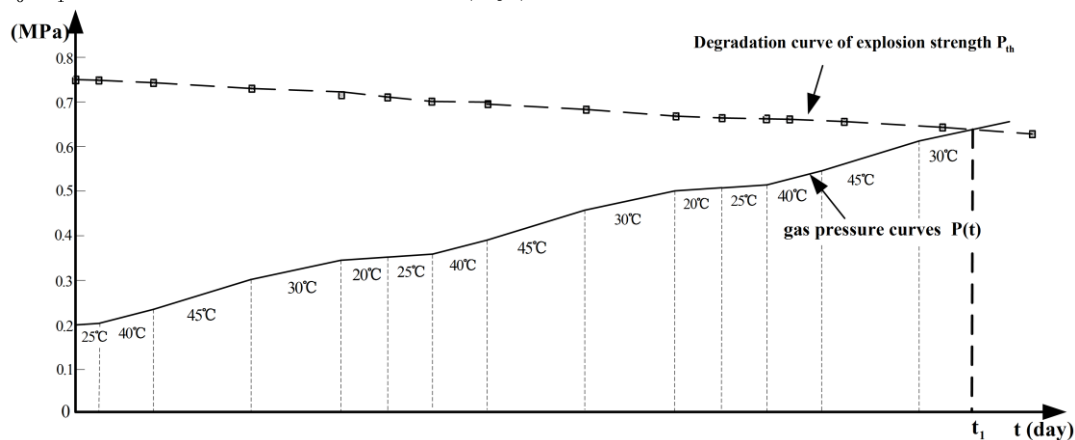


Figure 10 a degradation process of the explosion proof valve during the whole life period

Conclusion

In this paper, a two-phase prediction life model of explosion proof valve has been established. Based on ADT and explosion strength degradation model, the explosion safety valve life prediction is achieved, and main conclusions are drawn as follows. Test points at each stress are closely and uniformly distributed around the corresponding fitting curves, and the determination coefficient at each curve is much closer to 1. Which indicates the test plan is feasible and the high calculation accuracy of fitting formulas. It also reflects actual engineering practice in battery life prediction.

References

- Balakrishnan PG, Ramesh R and Prem Kumar T. (2006) Safety mechanisms in lithium-ion batteries. *Journal Of Power Sources* 155: 401-414.
- Belharouak I, Sun YK and Liu J, et al. (2003) Li(Ni1/3Co1/3Mn1/3)O2 as a suitable cathode for high power applications. *Journal Of Power Sources* 123: 247-252.
- Doksum KA and Hóyland A. (1992) Models for Variable-Stress Accelerated Life Testing Experiments Based on Wener Processes and the Inverse Gaussian Distribution. *Technometrics* 34: 74-82.

- Kim SK, Jeong WT and Lee HK, et al. (2008) Characteristics of $\text{LiNi}_{1/3}\text{Co}_{1/3}\text{Mn}_{1/3}\text{O}_2$ Cathode Powder Prepared by Different Method in Lithium Rechargeable Batteries. *International Journal of Electrochemical Science* 3: 1504-1511.
- Kumai K, Miyashiro H and Kobayashi Y, et al. (1999) Gas generation mechanism due to electrolyte decomposition in commercial lithium-ion cell. *Journal Of Power Sources* s 81–82: 715-719.
- Peng CY and Tseng ST. (2010) Progressive-Stress Accelerated Degradation Test for Highly-Reliable Products. *Reliability IEEE Transactions on* 59: 30-37.
- Séronie-Vivien S, Toullec S and Malard L, et al. (2012) Gas Desorption Behavior of Graphite Anode in Lithium Ion Secondary Battery and its Effect on the Battery Performance. *Journal of the Vacuum Society of Japan* 55: 395-398.
- Shaju KM and Bruce PG. (2006) Macroporous $\text{Li}(\text{Ni}_{1/3}\text{Co}_{1/3}\text{Mn}_{1/3})\text{O}_2$: A High-Power and High-Energy Cathode for Rechargeable Lithium Batteries. *Advanced Materials* 18: 2330-2334.
- Shuang L, Xiong L and He C. (2014) Long cycle life lithium ion battery with lithium nickel cobalt manganese oxide (NCM) cathode. *Journal Of Power Sources* 261: 285-291.
- Sun F, Markötter H and Manke I, et al. (2016) Three dimensional visualization of gas evolution and channel formation inside a lithium-ion battery. *ACS applied materials & interfaces*.

Maintenance optimization for systems with non-stationary degradation and random shocks

Chiming Guo, Yongsheng Bai, Yunxian Jia

Mechanical Engineering College, Shijiazhuang, China, 050003, email: guochiming@nurd.edu.cn

Abstract. Many systems suffer competing risks of degradation processes and random shocks, but much less efforts are devoted to consider the competing risks with the non-stationary deterioration process and random shocks. In this paper, an adaptive maintenance policy is proposed to deal with the competing risks with a non-stationary degradation process. Two optimization models, off-line and on-line, are used to evaluate the maintenance policy, respectively. Through a comparison with the blind maintenance optimization results, it shows that it is important to consider the change point and proves the effectiveness of the on-line optimization model.

1. Introduction

In reality, many systems suffer competing failure processes, that is, the system has multiple failure mechanisms at the same time and may fail due to one of them, such as wear, corrosion, shock loads, etc. Therefore, it is beneficial to consider the competing risks for the maintenance scheduling to enhance the system availability. Researches (Lehman 2009, Li and Pham 2005, Huynh et al. 2012, Peng et al 2011) have investigated the competing risks based on stationary degradation processes. However, the degradation processes for many systems are non-stationary due to environment influences or ageing factor etc. (Wang 2007, Jiang 2011). To identify the degradation change point is very significant for the maintenance planning. Some researchers have considered the change point detection in the CBM settings. Fouladirad and Grall (2011) proposed an adaptive PM threshold model for a system with a wear rate transition. The cumulative sum (CUSUM) algorithm is used to detect the abrupt change time. Ponchet et al. (2010) developed two condition-based maintenance optimization models with and without considering the abrupt changes in their degradation processes, respectively. The numerical results show that the change of degradation mode strongly influences the choice of the best decision rule structure. Therefore, it is necessary to consider the competing failures with degradation changes in the maintenance scheduling problems.

In this paper, a maintenance optimization model considering the competing risks caused by the non-stationary degradation process and shock processes will be proposed. The main contributions of this study are: (a) developing an analytical cost model for the competing risk with non-stationary degradation processes in the adaptive maintenance policies; (b) performing a comparison between the offline results with the online results in the maintenance optimization.

2. Reliability analysis with degradation and shocks

The failure causes of the considered system are competing due to degradation and random shocks. The system fails if the deterioration reaches a critical threshold or when a shock occurs.

2.1. Degradation model

The system is subjected to continuous accumulation of deterioration. The system degradation level at time t can be denoted by a random variable $X(t)$. $\{X(t)\}_{t \geq 0}$ is an increasing stochastic process with initial state

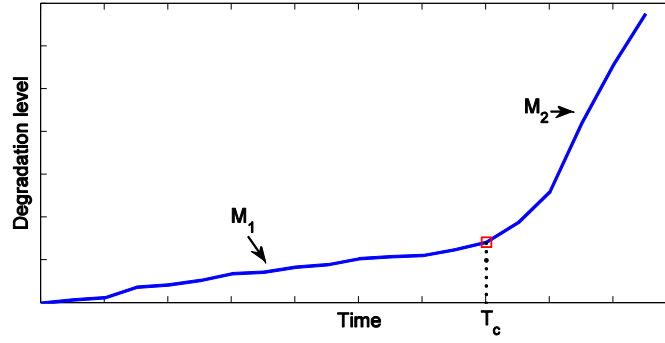


Figure 1. Degradation path with change point

$X(0)=0$. Define the failure time as $T_L := \inf\{t : X(t) \geq L, t \geq 0\}$, and L is the degradation failure threshold. The degradation failure does not mean that the system cannot be in operation, but means that its low reliability is unacceptable for economic or safety reasons. Due to the internal mechanism or the external environment influences, the deterioration rate can undergo a sudden change at a random time T_c during the system life cycle. The system degradation process is divided into nominal mode M_1 and accelerated mode M_2 , see Figure 1.

Assume the deteriorating process in mode M_i ($i=1,2$), denoted by $\{X_t^i\}_{t \geq 0}$ is a gamma process. The increment $Y_{t-s}^i = X_t^i - X_s^i$ between t and s ($0 \leq s \leq t$) follows a gamma probability density function (pdf) with shape parameter $\alpha_i(t-s)$ and scale parameter β_i :

$$f_{\alpha_i(t-s), \beta_i}(y) = \frac{\beta_i^{\alpha_i(t-s)}}{\Gamma(\alpha_i(t-s))} y^{\alpha_i(t-s)-1} e^{-\beta_i y} \mathbf{I}_{\{y \geq 0\}}, \quad (1)$$

where $\Gamma(\alpha_i(t-s)) = \int_0^\infty w^{\alpha_i(t-s)-1} e^{-w} dw$, $\alpha_i(t-s) > 0$, $\beta_i > 0$, $i=1,2$. $\mathbf{I}_{\{E\}} = 1$ if E is true and 0 otherwise.

The average of deterioration rate in M_i is $\mu_i = \alpha_i / \beta_i$ and the variance is $\sigma_i^2 = \alpha_i / \beta_i^2$. For the sudden change in the deterioration process, the mean deterioration rate in the accelerated mode M_2 is larger than the mean deterioration rate in the nominal mode M_1 , that is $\alpha_2 / \beta_2 > \alpha_1 / \beta_1$. According to the definition of the deterioration rate transition, the system state at time t can be represented as

$$X(t) = Y_t^1 \mathbf{I}_{\{t \leq T_c\}} + (Y_{T_c}^1 + Y_{t-T_c}^2) \mathbf{I}_{\{t > T_c\}}. \quad (2)$$

The increment $X(t) - X(s)$ depends not only on the time interval $t-s$ but also on the deterioration mode. The degradation process is a non-homogeneous gamma process.

2.2. Shock model

In many practical situations, the system is not only subject to degradation failure but also shock (or catastrophic) failure, due to external or internal shocks. Assume the shock failure time T_s follow Weibull distribution,

$$F_s(t) = 1 - \exp[-(t / \mu)^\theta]. \quad (3)$$

The shock failure rate is

$$\lambda_s(t) = \frac{\theta}{\mu} \left(\frac{t}{\mu} \right)^{\theta-1}. \quad (4)$$

It can be seen that the shock rate increases with time t , since the system is more vulnerable to the shocks when the system deteriorate seriously.

3. Online CUSUM algorithm

Online change detection uses the online information to detect the change time. CUSUM algorithm is one of the widely used change-point detection algorithms (Basseville and Nikiforov 1993). In this paper, we choose CUSUM algorithm as the detection procedure when there is no priori information about the change time.

Let $Y_1, Y_2, \dots, Y_v$ denote the subsequent deterioration increments in a fixed time interval Δt before T_c with the density function $f_{\alpha_1 \Delta t, \beta_1}(y_j)$, and $Y_{v+1}, Y_{v+2}, \dots$ denotes the subsequent deterioration increments in a fixed time interval Δt after T_c with density function $f_{\alpha_2 \Delta t, \beta_2}(y_j)$, where $v \in Z^+$ and y_j is the realization of Y_j .

Define $Q_n = \max_{1 \leq k \leq n} \sum_{j=k}^n \log\left(\frac{f_{\alpha_2 \Delta t, \beta_2}(y_j)}{f_{\alpha_1 \Delta t, \beta_1}(y_j)}\right)$, $Q_0 = 0$. According to the one-sided CUSUM scheme, the stopping rule for the sequential observations $\{Y_j\}_{1 \leq j \leq n}$ is

$$N = \min\{n : \max_{1 \leq k \leq n} \sum_{j=k}^n \log \frac{f_{\alpha_2 \Delta t, \beta_2}(y_j)}{f_{\alpha_1 \Delta t, \beta_1}(y_j)} \geq c_\gamma\}. \quad (5)$$

When $Q_n \geq c_\gamma$, the system will be judged in M_2 . Then the detection time of the system mode change is $t_{\text{detect}} = N\Delta t$.

4. Maintenance policies

There are three possible maintenance actions, inspection, preventive replacement and corrective replacement, respectively. The preventive/corrective replacement restores the system to be as good as new with negligible time. For the cost reason, the system is periodically inspected without error. The system degradation status can only be revealed by inspection. The inspection times $\{t_k\}_{k \in \mathbb{N}}$ defined by $t_k = k\Delta t$.

According to whether considering the change point in the degradation process, two maintenance policies are considered in this paper: adaptive maintenance policy and change-blind maintenance policy.

4.1 Adaptive maintenance policy

The preventive maintenance threshold varies for different degradation modes in the adaptive maintenance policy. The detailed adaptive maintenance policy is as follows.

- If one of the following exclusive events $\{X(t_{k-1}) < A_1 \cap X(t_k) \geq L \cap t_{k-1} < t_k \leq t_c\}$, $\{X(t_{k-1}) < A_2 \cap X(t_k) \geq L \cap t_c \leq t_{k-1} < t_k\}$ or $\{X(t_{k-1}) < A_1 \cap X(t_k) \geq L \cap t_{k-1} \leq t_c < t_k\}$ occurs before a shock failure, a degradation failure happens and a corrective replacement is carried out at time t_k with cost C_C . t_c is the realization of T_c .
- If one of the following exclusive events $\{X(t_{k-1}) < A_2 \cap t_c \leq t_{k-1}\}$, $\{X(t_{k-1}) < A_1 \cap t_{k-1} \leq t_c < t_k\}$ or $\{X(t_{k-1}) < A_1 \cap t_k \leq t_c\}$ occurs and a shock failure happens at time t , ($t_{k-1} < t \leq t_k$), then a corrective replacement is carried out immediately with cost C_C .
- If one of the following exclusive events $\{X(t_{k-1}) < A_1 \cap A_1 \leq X(t_k) < L \cap t_{k-1} < t_k \leq t_c\}$, $\{X(t_{k-1}) < A_1 \cap A_2 \leq X(t_k) < L \cap t_{k-1} \leq t_c < t_k\}$ or $\{X(t_{k-1}) < A_2 \cap A_2 \leq X(t_k) < L \cap t_c \leq t_{k-1} < t_k\}$ occurs before a shock failure, a preventive replacement is triggered at time t_k with cost C_P .
- If one of the following exclusive events $\{X(t_k) < A_1 \cap t_k \leq t_c\}$ or $\{X(t_k) < A_2 \cap t_k \geq t_c\}$ occurs before a shock failure, the system is left unchanged, and the maintenance decision is postponed to the next inspection time t_{k+1} . Each inspection incurs a cost C_I .
- If the degradation failure happens between two inspections, there will be a period of unavailability for the system, and the per unit time cost for the unavailability is C_D .

Considering the change-point effect, the maintenance policy has two preventive replacement thresholds, A_1 and A_2 for different degradation modes, respectively. The threshold A_1 can be obtained by minimizing the maintenance cost in mode M_1 and A_2 can be obtained in mode M_2 by the same way. For mode M_2 is an accelerated degradation mode, we have $A_2 < A_1$.

4.2 Change-blind maintenance policy

The decision rule of the change-blind maintenance policy neglects the change of degradation rate and only considers the system stay in mode M_1 . The system is inspected periodically with an inspection interval Δt . A_1 and Δt are two maintenance decision variables in the change-blind maintenance policy. The corresponding costs for different maintenance actions are the same as the adaptive maintenance policy.

5. Adaptive maintenance policy evaluation

Considering the cost incurred for each maintenance action, the maintenance policy is evaluated using the expected long run cost rate over an infinite time span. Because the system is restored to as good as new after each preventive/corrective replacement, we can use the renewal reward theory (Ross 1996) to compute the expected long-run cost rate as

$$E[C_\infty] = \lim_{t \rightarrow \infty} \frac{E[C(t)]}{t} = \frac{E[C(T)]}{E[T]}, \quad (6)$$

where $C(t)$ is the total cost at time t , T is the length of a renewal cycle.

The accumulated cost on a renewal cycle T can be written as,

$$E[C(T)] = C_I E[N_I(T)] + C_P P_P(\Delta t, \mathbf{A}) + C_C P_C(\Delta t, \mathbf{A}) + C_D E[W(T)], \quad (7)$$

where

$E[N_I(T)]$ is the expected number of inspections during a renewal cycle T ;

$P_P(\Delta t, \mathbf{A})$ is the probability that the renewal cycle ends with a preventive replacement and $\mathbf{A}$ is the PM threshold set $\{A_1, A_2\}$;

$P_C(\Delta t, \mathbf{A})$ is probability that the renewal cycle ends with a corrective replacement;

$E[W(T)]$ is the expected unavailability time in a renewal cycle.

The adaptive maintenance optimization problem is reduced to find the value of Δt , A_1 and A_2 that minimize the expected long-run cost rate $E[C_\infty]$.

According to whether the distribution of the change point is known, two configurations for the adaptive maintenance policy evaluation are considered: online evaluation and offline evaluation.

5.1. Online evaluation

When the change time is completely unknown, CUSUM algorithm is used to detect the change point in the degradation process and there will be a detection delay for the change time. In this configuration, the expected long-run cost rate is evaluated based on the Monte Carlo simulation with online CUSUM algorithm for different inspection interval.

5.2. Offline evaluation

If the distribution of the change time is known before the application of the adaptive maintenance, the analytical expression of the expected long-run cost rate can be obtained. There will be no detection delay of the change time in this case. The detailed analytical expressions are as follows.

5.2.1. Expression of $P_C(\Delta t, \mathbf{A})$

The corrective replacement can be caused by degradation failure or shock failure, so the probability for a corrective replacement in a renewal cycle is denoted as

$$P_C(\Delta t, \mathbf{A}) = P_{DC}(\Delta t, \mathbf{A}) + P_{SC}(\Delta t, \mathbf{A}), \quad (8)$$

where $P_{DC}(\Delta t, \mathbf{A})$ is the probability for a corrective replacement due to a degradation failure, $P_{SC}(\Delta t, \mathbf{A})$ is the probability for a corrective replacement due to a shock failure.

Through classifying the change point occasions (see Figure.2), the probability for the degradation failure before a shock failure can be expressed as

$$P_{DC}(\Delta t, \mathbf{A}) = \sum_{k=1}^{\infty} \left(\begin{aligned} &P(X(t_{k-1}) < A_1 \cap X(t_k) \geq L \cap t_{k-1} < t_k \leq t_c) \\ &+ P(X(t_{k-1}) < A_2 \cap X(t_k) \geq L \cap t_c \leq t_{k-1} < t_k) \\ &+ P(X(t_{k-1}) < A_1 \cap X(t_k) \geq L \cap t_{k-1} \leq t_c < t_k) \end{aligned} \right) R_S(t_k), \quad (9)$$

where $R_S(t_k) = \exp[-(t_k / \mu)^\theta]$ denotes the survival function of the shock process at t_k . The detail expression is given in Appendix A.

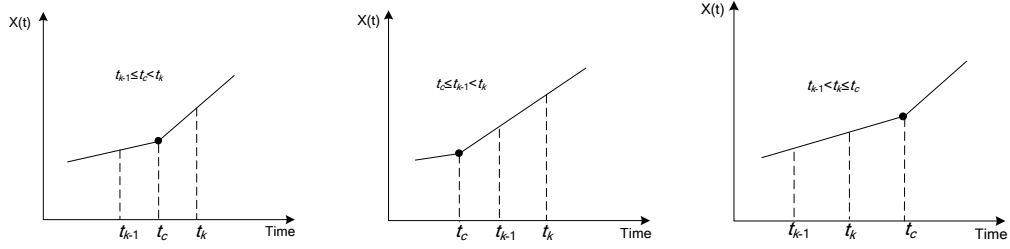


Figure 2. Different locations of t_c

Under the considered maintenance policy, the probability for the shock failure before a preventive replacement is expressed as

$$\begin{aligned}
 P_{SC}(\Delta t, \mathbf{A}) &= \sum_{k=1}^{\infty} \left(P(X(t_{k-1}) < A_2 \cap t_c \leq t_{k-1} \cap t_{k-1} < T_S < t_k) \right. \\
 &\quad \left. + P(X(t_{k-1}) < A_1 \cap t_{k-1} < t_c \cap t_{k-1} < T_S < t_k) \right) \\
 &= \sum_{k=1}^{\infty} (P(X(t_{k-1}) < A_2 \cap t_c \leq t_{k-1}) + P(X(t_{k-1}) < A_1 \cap t_{k-1} \leq t_c)) (R_S(t_{k-1}) - R_S(t_k)) R_S(t_{k-1}) \\
 &= \sum_{k=1}^{\infty} \left(\int_0^{t_{k-1}} \int_0^{A_2} \int_0^z f_{\alpha_1 t_c, \beta_1}(u) f_{\alpha_2(t_{k-1}-t_c), \beta_2}(z-u) f_c(t_c) du dz dt_c + \int_0^{A_1} \int_{t_{k-1}}^{\infty} f_{\alpha_1 t_{k-1}, \beta_1}(u) f_c(t_c) dt_c du \right) (R_S(t_{k-1}) - R_S(t_k)) R_S(t_{k-1}).
 \end{aligned} \tag{10}$$

5.2.2. Expression of $P_p(\Delta t, \mathbf{A})$

Under the considered maintenance policy, the probability for a renewal cycle ended by a preventive replacement is

$$P_p(\Delta t, \mathbf{A}) = \sum_{k=1}^{\infty} \left(P(X(t_{k-1}) < A_1 \cap A_1 \leq X(t_k) < L \cap t_{k-1} < t_k \leq t_c) \right. \\
 \left. + P(X(t_{k-1}) < A_2 \cap A_2 \leq X(t_k) < L \cap t_c \leq t_{k-1} < t_k) \right. \\
 \left. + P(X(t_{k-1}) < A_1 \cap A_2 \leq X(t_k) < L \cap t_{k-1} \leq t_c < t_k) \right) R_S(t_k). \tag{11}$$

where

$$\begin{aligned}
 &P(X(t_{k-1}) < A_1 \cap A_1 \leq X(t_k) < L \cap t_{k-1} < t_k \leq t_c) \\
 &= \int_{t_k}^{\infty} f_c(t_c) dt_c \int_0^{A_1} \int_{A_1-u}^{L-u} f_{\alpha_1 \Delta t, \beta_1}(\omega) f_{\alpha_1 t_{k-1}, \beta_1}(u) d\omega du
 \end{aligned} \tag{12}$$

$$\begin{aligned}
 &P(X(t_{k-1}) < A_2 \cap A_2 \leq X(t_k) < L \cap t_c \leq t_{k-1} < t_k) \\
 &= \int_0^{t_{k-1}} \int_0^{A_2} \int_0^z \int_{A_2-z}^{L-z} f_{\alpha_1 t_c, \beta_1}(u) f_{\alpha_2(t_{k-1}-t_c), \beta_2}(z-u) f_{\alpha_2 \Delta t, \beta_2}(\omega) f_c(t_c) d\omega du dz dt_c
 \end{aligned} \tag{13}$$

$$\begin{aligned}
 &P(X(t_{k-1}) < A_1 \cap A_2 \leq X(t_k) < L \cap t_{k-1} \leq t_c < t_k) \\
 &= \int_{t_{k-1}}^{t_k} f_c(t_c) dt_c \int_0^{A_1} \int_{\max(0, A_2-u)}^{L-u} \int_0^z f_{\alpha_1(t_c-t_{k-1}), \beta_1}(\omega) f_{\alpha_2(t_k-t_c), \beta_2}(z-\omega) f_{\alpha_1 t_{k-1}, \beta_1}(u) d\omega dz du
 \end{aligned} \tag{14}$$

5.2.3. Expression of $E[N_I(T)]$

Based on the formulas of the probability for the corrective/preventive replacement, the expected number of inspection during a renewal cycle $E[N_I(T)]$ is given as

$$E[N_I(T)] = \sum_{k=1}^{\infty} (\Lambda_P(k) + \Lambda_F(k)) R_S(t_k) k + \sum_{k=1}^{\infty} \Lambda_S(k) (k-1) \tag{15}$$

Where $\Lambda_P(k)$: the probability for the preventive replacement at the k th inspection; $\Lambda_F(k)$: the probability for degradation failure replacement at the k th inspection; $\Lambda_S(k)$: the probability for the shock failure before the k th inspection.

5.2.4. Expression of $E[W(T)]$

Under the considered maintenance policy, the degradation failure is non-self-announcing, and there will be a period of unavailability after degradation failure for the system. The unavailability time caused by the degradation failure can appear in the corrective replacement cycle with degradation failure or shock failure. Denote the expected unavailability time in a renewal cycle ended with a degradation failure (resp. shock failure) by $E[W_1(T)]$ (resp. $E[W_2(T)]$), then

$$E[W(T)] = E[W_1(T)] + E[W_2(T)] \quad (16)$$

The detail expression is given in Appendix B.

5.2.5. Expression of $E[T]$

Under the considered maintenance policy, the system lifecycle can be ended by a preventive replacement and a corrective replacement. The expected length of the system lifecycle is given as

$$E[T] = \sum_{k=1}^{\infty} (\Lambda_p(k) + \Lambda_F(k)) R_s(t_k) t_k + \sum_{k=1}^{\infty} E[T_{SC} | k] \quad (17)$$

where $E[T_{SC} | k]$ is the expected length of a renewal cycle ended with a shock failure before the k th inspection. The detail expression is given in Appendix C.

6. Numerical examples

In this section, the optimal maintenance decision values of Δt , A_1 and A_2 are investigated. The degradation model parameters are $(\alpha_1, \beta_1) = (1, 1)$, and $(\alpha_2, \beta_2) = (1, 0.2)$. The system degradation failure threshold $L = 50$. The shock model related parameters are $\mu = 60$, $\theta = 2$. The maintenance costs are, respectively, $C_i = 1$, $C_p = 5$, $C_c = 20$, $C_d = 50$. The change time T_c is assumed to follow uniform distribution $U(10, 20)$.

Considering the system with single degradation mode M_1 and shock failure, the optimal expected cost rate is computed by simulation as $E[C_{\infty}^1] = 0.46$, which is achieved at $A_1 = 2$, $\Delta t_1 = 26$. For the system with single degradation mode M_2 and shock failure, the optimal expected cost rate is computed by simulation as $E[C_{\infty}^2] = 1.53$, which is achieved at $A_2 = 23$, $\Delta t_2 = 3$. As $\Delta t_2 < \Delta t_1$, Δt_2 is chosen as the CUSUM algorithm detection interval. Therefore, The minimum cost rate with mode M_1 and shock failure is recalculated at $\Delta t_1 = 3$, the optimal results are $E[C_{\infty}^1] = 0.76$ with $A_1 = 25$, $\Delta t_1 = 3$.

6.1. Adaptive maintenance policy

Online optimization: When $\Delta t = 3$, $A_1 = 25$, $A_2 = 23$, the minimum $E[C_{\infty}] = 0.87$ is achieved by $h = 95$ with CUSUM algorithm.

Offline optimization: When $\Delta t = 3$, $A_1 = 25$, $A_2 = 23$, the minimum $E[C_{\infty}] = 0.86$ is achieved by Monte Carlo simulation.

Compared with the two evaluation results, it finds that the achieved optimal maintenance cost rate by offline model is smaller than the optimal result of online model. The offline model uses the change point time distribution and has no change point detection delay time, however, the online model without considering the change point time distribution has change point detection delay. The delay time makes the optimal cost rate of online model larger than the offline model. Meanwhile, the online optimization result is similar with the offline optimization result, and it proves that online adaptive maintenance is suitable for the system with change point.

6.2. Change-blind maintenance policy

The blind maintenance policy neglects the change points and considers the system always stay in degradation mode M_1 . Through the optimal results $A_1 = 2$, $\Delta t_1 = 26$ for the system with single degradation mode M_1 and shock failure, the minimum expected cost rate is calculated as $E[C_{\infty}] = 2.01$. Compared with the adaptive maintenance evaluation results, the maintenance cost of change-blind maintenance policy is larger. Because change-blind maintenance policy doesn't consider the dynamic changes of the system and cause the system failure cost increase.

7. Conclusions

For non-stationary degradation, this paper develops adaptive maintenance optimization models under the competing risk failures with degradation failure and sudden failure. According to whether the change point time distribution is known, the adaptive maintenance evaluation is divided into online evaluation and offline evaluation. Through classifying the change point occasions, and the analytical model for the offline evaluation is developed. The results prove that the online maintenance can provide an effective adaptive maintenance. It is necessary to consider the system dynamic changes in the maintenance decision process.

Acknowledgements

This research was supported by the National Science Foundation of China under agreement 71401173.

Appendix A

$$\begin{aligned} P(X(t_{k-1}) < A_1 \cap X(t_k) \geq L \cap t_{k-1} < t_k \leq t_c) \\ = \int_{t_k}^{\infty} \int_0^{A_1} \int_{L-z}^{\infty} f_{\alpha_1 \Delta t, \beta_1}(u) f_{\alpha_1 t_{k-1}, \beta_1}(z) f_c(t_c) du dz dt_c \end{aligned} \quad (A.1)$$

$$\begin{aligned} P(X(t_{k-1}) < A_2 \cap X(t_k) \geq L \cap t_c \leq t_{k-1} < t_k) \\ = \int_0^{t_{k-1}} \int_0^{A_2} \int_0^z \int_{L-z}^{\infty} f_{\alpha_2 \Delta t, \beta_2}(\omega) f_{\alpha_1 t_c, \beta_1}(u) f_{\alpha_2(t_{k-1}-t_c), \beta_2}(z-u) f_c(t_c) d\omega du dz dt_c \end{aligned} \quad (A.2)$$

$$\begin{aligned} P(X(t_{k-1}) < A_1 \cap X(t_k) \geq L \cap t_{k-1} \leq t_c < t_k) \\ = \int_{t_{k-1}}^{t_k} \int_0^{A_1} \int_0^{\infty} \int_0^z f_{\alpha_1(t_c-t_{k-1}), \beta_1}(\omega) f_{\alpha_2(t_k-t_c), \beta_2}(z-\omega) f_{\alpha_1 t_{k-1}, \beta_1}(u) f_c(t_c) d\omega dz du dt_c \end{aligned} \quad (A.3)$$

Appendix B

The expected unavailability time in a renewal cycle ended with a degradation failure, $E[W_1(T)]$ is given as

$$E[W_1(T)] = \sum_{k=1}^{\infty} (E[W_1(T, k)]) \quad (B.1)$$

$$\begin{aligned} E[W_1(T, k)] &= \underbrace{\int_0^{A_1} \int_0^{\Delta t} (\Delta t - \omega) f_{L-u}(\omega | \alpha_1, \beta_1) f_{\alpha_1 t_{k-1}, \beta_1}(u) d\omega du \int_{t_k}^{\infty} f_c(t_c) dt_c}_{t_{k-1} < t_k \leq t_c} \\ &+ \underbrace{\int_0^{t_{k-1}} \int_0^{A_2} \int_0^z \int_0^{\Delta t} (\Delta t - \omega) f_{L-z}(\omega | \alpha_2, \beta_2) f_{\alpha_1 t_c, \beta_1}(u) f_{\alpha_2(t_{k-1}-t_c), \beta_2}(z-u) f_c(t_c) d\omega du dz dt_c}_{t_c \leq t_{k-1} < t_k} \\ &+ \underbrace{\int_{t_{k-1}}^{t_k} \int_0^{A_1} \int_0^{\Delta t} (\Delta t - \omega) f_{L-u}(\omega | \alpha_1, \beta_1) f_{\alpha_1 t_{k-1}, \beta_1}(u) f_c(t_c) d\omega du dt_c}_{t_{k-1} < T_L \leq t_c < t_k} \\ &+ \underbrace{\int_{t_{k-1}}^{t_k} \int_0^{A_1} \int_0^{\Delta t} (\Delta t - \omega) f_{L-u}(\omega | \alpha_1, \beta_1, \alpha_2, \beta_2) f_{\alpha_1 t_{k-1}, \beta_1}(u) f_c(t_c) d\omega du dt_c}_{t_{k-1} < t_c < T_L \leq t_k}, \end{aligned} \quad (B.2)$$

where

$$f_{L-u}(t | \alpha_1, \beta_1) = \frac{\partial F_{L-u}(t | \alpha_1, \beta_1)}{\partial t}, \quad F_{L-u}(t | \alpha_1, \beta_1) = \int_{L-u}^{\infty} f_{\alpha_1 t, \beta_1}(\omega) d\omega,$$

$$f_{L-z}(t | \alpha_2, \beta_2) = \frac{\partial F_{L-z}(t | \alpha_2, \beta_2)}{\partial t}, \quad F_{L-z}(t | \alpha_2, \beta_2) = \int_{L-z}^{\infty} f_{\alpha_2 t, \beta_2}(\omega) d\omega,$$

$$F_{L-u}(t | \alpha_1, \beta_1, \alpha_2, \beta_2) = \int_{L-u}^{\infty} \int_0^z f_{\alpha_1(t-t_{k-1}), \beta_1}(\omega) f_{\alpha_2(t-t_c), \beta_2}(z-\omega) d\omega dz,$$

$$f_{L-u}(\omega | \alpha_1, \beta_1, \alpha_2, \beta_2) = \frac{\partial F_{L-u}(t | \alpha_1, \beta_1, \alpha_2, \beta_2)}{\partial t}.$$

An extended log-linear model for modeling wear processes

R. Jiang*

*Faculty of Automotive and Mechanical Engineering Changsha University of Science and Technology, Changsha, Hunan 410114, China. E-mail address: jiang@csust.edu.cn.

Abstract. The purpose of this paper is twofold. First, we show that a bounded process model (e.g., log-linear process) can be inappropriate for describing a cumulative degradation amount (e.g., accumulated wear amount) since the bounded process can result in an invalid distribution of time to failure. Then, we use an extended log-linear process model to describe the mean value function of a non-homogeneous degradation process. The extended model can be viewed as a special case of the model superposed by log-linear process and power law process, and is no longer bounded. Its intensity function can be constant, increasing or decreasing. A real-world example that deals with a non-homogeneous wear process is used to illustrate the above conclusion and the appropriateness of the extended model. For the example under consideration, the Akaike information criterion supports the use of log-linear process though the reliability function derived from it is not a valid model. This implies that a model that offers a good fit to the observed degradation process is not necessary an appropriate model. Applications of the resulting model for the numerical example are also discussed.

1. Introduction

The wear process is a typical degradation process and can be modeled using stochastic processes. The system is regarded as failed when the degradation process reaches a critical threshold (e.g., wear limit in the case of wear processes). Bocchetti et al. (2009) use the non-homogeneous Poisson process (NHPP) to model the wear degradation for the cylinder liners of a marine diesel engine and use the log-linear process (LLP) to describe the expected number of wear events. The cumulative wear amount described by LLP is bounded. We show that the models with a bounded cumulative degradation amount are not a valid model since the reliability function derived from such models never tends to zero.

The paper presents an extended LLP, which is a valid model for representing the expected cumulative wear amount of a wear process. The new model can represent three different degradation patterns:

- (a) the degradation intensity is constant,
- (b) the degradation intensity is increasing and tends to a constant, and
- (c) the degradation intensity is decreasing and tends to a constant.

The non-homogeneous gamma process is used to model the degradation increment and to derive the distribution of time to the wear limit. The model parameters are estimated using the maximum likelihood method (MLM).

The proposed model is applied to fit a set of real-world data. The resulting model is applied to predict the distribution of time to the wear limit and to make condition-based maintenance decision. The prediction and decision processes are illustrated for three specific degradation processes.

The paper is organized as follows. The LLP for modeling the expected number of cumulative wear events is analyzed in Section 2. The proposed model is presented in Section 3. Prediction of time to the wear limit for the population and individuals are discussed in Section 4. The paper is concluded in Section 5.

2. Wear data and LLP model

2.1 Wear data

The data shown in Table 1 come from Bocchetti et al. (2009), and deal with accumulated wear amounts of 33 cylinder liners of 8-cylinder SULZER RTA 58 engines. In the table, t_{ij} (in 1000 hours) is the j th inspection time or operating age ($j = 1, 2, \dots, n_i$) for the i th liner ($i = 1, 2, \dots, 33$), and w_{ij} is the corresponding accumulated wear amount in mm. There is no wear record for the 32nd liner, and the 12th, 25th and 28th liners were replaced at the last inspection due to a large accumulated wear. The maximum admissible wear is 4 mm.

Table 1. Wear data of cylinder liners

i	t_{i1}	w_{i1}	t_{i2}	w_{i2}	t_{i3}	w_{i3}	t_{i4}	w_{i4}
1	11.3	0.90	14.68	1.30	31.27	2.85		
2	11.36	0.80	17.2	1.35				
3	11.3	1.50	21.97	2.00				
4	12.3	1.00	16.3	1.35				
5	14.81	1.90	18.7	2.25	28	2.75		
6	9.7	1.10	19.71	2.60	30.45	3.00		
7	10	1.20	30.45	2.75	37.31	3.05		
8	6.86	0.50	17.2	1.45	24.71	2.15		
9	2.04	0.40	12.58	2.00	16.62	2.35		
10	7.54	0.50	8.84	1.10	9.77	1.15	16.3	2.10
11	8.51	0.80	14.93	1.45	21.56	1.90		
12	18.32	2.20	25.31	3.00	37.31	3.70	45	3.95
13	10	2.10	16.62	2.75	30	3.60		
14	9.35	0.85	15.97	1.20				
15	13.2	2.0						
16	7.7	1.05						
17	7.7	1.60						
18	8.25	0.90						
19	3.9	1.15						
20	7.7	1.20						
21	9.096	0.50						
22	19.8	1.60						
23	10.45	0.40						
24	12.1	1.00						
25	12	1.95	27.3	2.70	49.5	3.15	56.12	4.05
26	8.8	1.40						
27	2.2	0.40						
28	33	2.90	38.5	3.25	55.46	4.10		
29	8.8	0.50	27.5	2.15				
30	8.25	0.70						
31	18.755	1.15						
32								
33	8.49	0.95						

2.2 Inappropriateness of the log-linear process model

Bocchetti et al. (2009) assume:

- The wear of a cylinder liner at a given age t is the accumulation of isolated damage (elementary wear) of equal magnitude c that has occurred by age t . This assumption implies that the wear process is a jump process, and the possible values are $0, c, 2c, \dots$
- The events that cause an elementary wear occur randomly in time; and
- The probability that damage occurs in a given time interval is independent of the previously accumulated damage.

It is further assumed that the wear process cannot grow unlimitedly. The wearing process is described through a non-homogeneous Poisson process (NHPP) with intensity function being a negative exponential function (often referred to as log-linear process) given by

$$m(t) = ae^{-bt}, \quad a, b > 0. \quad (1)$$

The expected number of wear events up to the operating time t is given by

$$M(t) = E[N(t)] = \frac{a}{b}(1 - e^{-bt}). \quad (2)$$

It is noted that $M(\infty) = a/b$. That is, the expected number of wearing events is bounded and this may cause a big problem if $M(\infty)$ is not large enough as shown later.

The wear accumulated at the operating time t is modeled by the stochastic process

$$W(t) = cN(t). \quad (3)$$

Let $\Delta W(t_1, t_2) = c\Delta N(t_1, t_2) = c[N(t_2) - N(t_1)]$ denote the wear accumulated in the time interval (t_1, t_2) . Its probability distribution is given by

$$\Pr[\Delta W(t_1, t_2) = w] = \Pr[\Delta N(t_1, t_2) = w/c] = \frac{[\Delta M(t_1, t_2)]^n}{n!} e^{-\Delta M(t_1, t_2)} \quad (4)$$

where $n = w/c$, and $\Delta M(t_1, t_2) = M(t_2) - M(t_1)$.

Let L denote the wear limit and $K = \text{int}[L/c]$, where $\text{int}(x)$ rounds x down to a nearest integer. The reliability is the probability of the event that $W(t)$ does not exceed the wear limit, given by

$$R(t) = \sum_{i=0}^K \frac{M^i(t)}{i!} e^{-M(t)}. \quad (6)$$

As a valid reliability function, it requires that $R(\infty) = 0$. This is equivalent to $M(\infty) = \infty$. That is, the wear process must be unbounded (at least it must be a value that is much larger than K). As a result, the LLP model used for fitting the wear data is inappropriate.

To illustrate the above analysis, we look at the fitted model. Its model parameters are:

$$a = 0.7101, b = 0.02423, c = 0.1810, K = 22 \text{ and } a/b = 29.31.$$

Using these parameter values to Eq. (6), we obtained the plot of $R(t)$ shown in Figure 1. As seen, $R(t)$ tends to a positive constant (i.e., 0.1011) rather than zero. This is because the possibly largest expected number of wear events is 29, which is not much larger than K ($= 22$).

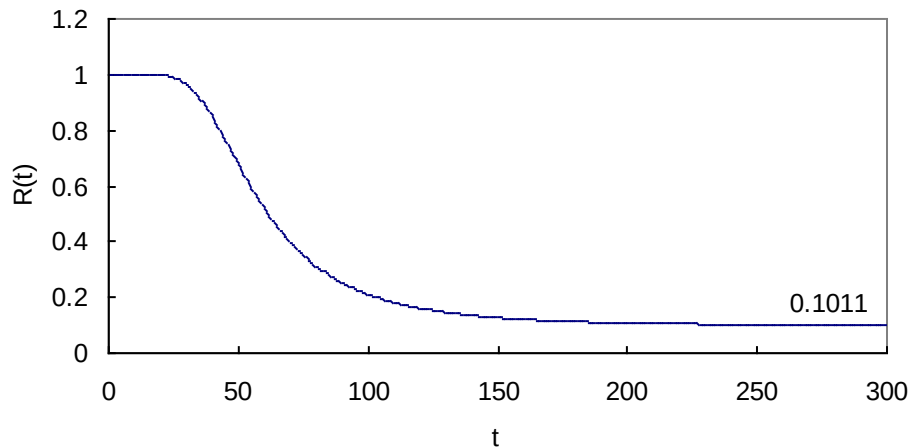


Figure 1. Plot of $R(t)$ derived from Eq. (6)

3. Proposed model and parameter estimation

3.1 Proposed model

Eq. (1) assumes that the rate of wear decreases to zero. This is unreasonable and will underestimate the failure risk. A more reasonable assumption is that the rate of wear decreases and tends to a positive constant. Guida et al (2014) mention this “conservative hypothesis”. According to this assumption, we have the extended LLP, given by

$$dw(t)/dt = a_0 + a_1 e^{-t/\eta}, \quad a_0 \geq 0, \eta > 0 \quad (7)$$

where $w(t) = E[W(t)]$ is the mean value function of random wear $W(t)$. From Eq. (7), we have

$$w(t) = a_0 t + a_1 \eta (1 - e^{-t/\eta}). \quad (8)$$

It is noted that the model reduces into the LLP model when $a_0 = 0$ and $a_1 > 0$.

Three possible cases for Eq. (7) are as follows:

- when $a_0 > 0$ and $-a_0 < a_1 < 0$, the rate of wear increases and tends to a positive constant,
- when $a_0 > 0$ and $a_1 = 0$, the rate of wear is a positive constant, and
- when $a_0, a_1 > 0$, the rate of wear decreases and tends to a positive constant.

3.2 Parameter estimation

Consider a wear path given by $(t_{ij}, w_{ij}; j = 0, 1, 2, \dots, n_i)$ and assume that $t_{i0} = 0$ and $w_{i0} = 0$. The wear increment $\Delta W_{ij} = W_{ij} - W_{i,j-1}$ is a random variable with mean given by

$$\mu_{ij} = E(\Delta W_{ij}) = a_0 \Delta t_{ij} + a_1 (e^{-t_{i,j-1}/\eta} - e^{-t_{ij}/\eta}), \Delta t_{ij} = t_{ij} - t_{i,j-1}. \quad (9)$$

We assume that the wear increment ΔW_{ij} is independent and follows the gamma distribution with shape parameter $u_{ij} = \mu_{ij} / v$ and the scale parameter v . Since μ_{ij} is a non-linear function of t , the degradation increment model is a non-homogeneous gamma process model. The likelihood function for the path i is given by

$$L_i = \prod_{j=1}^{n_i} g(\Delta w_{ij}; u_{ij}, v), \Delta w_{ij} = w_{ij} - w_{i,j-1}. \quad (10)$$

The overall log-likelihood function is given by

$$\ln(L) = \sum_{i=1}^n \ln(L_i). \quad (11)$$

The maximum likelihood estimates of the model parameters are shown in the second row of Table 2; and the maximum likelihood estimates of parameters of the original model are shown in the last row. In terms of Akaike information criterion, the original LLP model seemingly outperforms the extended model. However, the original model is not a valid reliability model since it gives $K = 26$, $a/b = 34.36$ and $R(\infty) = 0.0857 > 0$. As such, the proposed model is more appropriate.

Table 2. Estimates of the model parameters

Model	a_0	a_1	η	v	$\ln(L)$
Extended	0.0341	2.1944	21.3963	0.1483	-22.4445
Original	0	5.1356	39.6872	0.1495	-22.7895

4. Prediction of time to the wear limit for population and individuals

4.1 Distribution of time to the wear limit for the population

The cdf of time to the wear limit ($L = 4$ mm) is given by

$$F(t) = 1 - G(L; w(t) / v, v) \quad (12)$$

where $G(x; p, q)$ is the gamma cdf with the shape parameter p and scale parameter q , and $w(t)$ is given by Eq. (8). It is noted that $F(t)$ is a complex function of t . We use a numerical method to evaluate the density function. For a small value of Δt , the density function can be evaluated by

$$f(t + 0.5\Delta t) \approx [F(t + \Delta t) - F(t)] / \Delta t. \quad (13)$$

For the current example, the cdf evaluated by Eq. (13) is shown in Figure 2 (the solid line), and can be approximated by the gamma distribution $G(t; 10.14, 5.96)$ (see the dotted line in Figure 2).

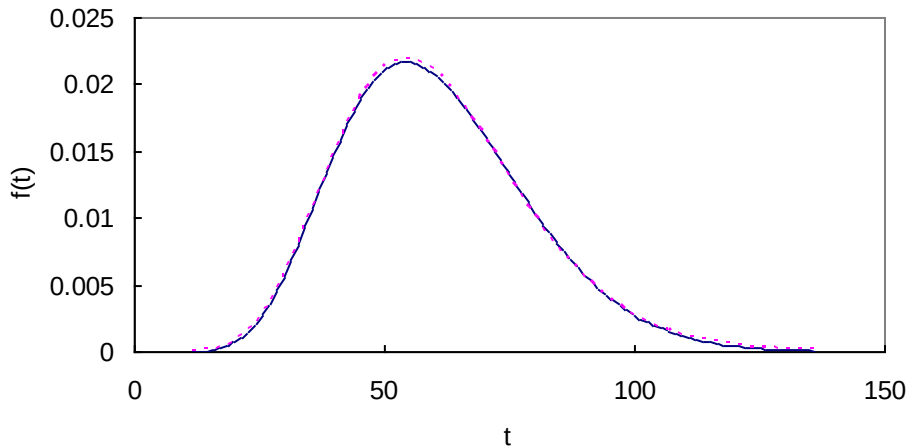


Figure 2. Distribution of time to the wear limit for the population

A possible application of the population distribution is to determine a preventive replacement age for the population. This needs a decision model such as the cost model (when the costs for a preventive replacement and a failure replacement are known) or the tradeoff BX life B_X^* (when the replacement costs are unknown, see Jiang 2013). For the current example, $B_X^* = 44.06$ (1000 hours).

For Liners 12, 25 and 28, the times to the wear limit can be inferred using linear interpolation or extrapolation based on the last two records and the results are shown in the second row of Table 3. As seen from the table, these liners would have been preventively replaced at B_X^* if the age policy with the preventive replacement age B_X^* were performed.

Table 3. Linearly inferred times to the wear limit T_f and maintenance types

	Liner 12	Liner 25	Liner 28
Linearly inferred T_f	46.54	55.75	53.46
Actual maintenance type	PM	CM	CM
Maintenance type based on B_X^*	PM	PM	PM

4.2 Distribution of time to the wear limit for Liners 12, 25 and 28

In this subsection we apply the resulting degradation model to make the maintenance decision for Liners 12, 25 and 28. The decision times t_d are shown in the second row of Table 4. The times for the mean wear curves to reach the wear limit are shown in the fourth row. As seen, the predicted failure times are close to the actual failure times for Liners 12 and 28 but much larger than the actual failure times for Liner 25.

Table 4. Maintenance decisions for Liners 12, 25 and 28

	Liner 12	Liner 25	Liner 28
Decision time t_d	37.31	49.50	38.50
Linearly inferred T_f	46.54	55.75	53.46
Predicted T_f	43.34	70.43	54.80
B_X^*	37.81	58.73	45.35
$B_X^* - t_d$	0.50	9.23	6.85
Maintenance decision	PM	Inspection at 54.50	Inspection at 43.50

However, the maintenance decision is not based on the predicted failure time; instead, it is based on the decision model (e.g., tradeoff BX life in this paper) and decision rules. A decision rule for the current example can be defined as: the next inspection will be scheduled if the time to the predicted preventive maintenance (PM) time, say $B_X^* - t_d$, is larger than the pre-specified inspection interval (e.g., 5000 hours), which can be optimized (e.g., see Jiang 2010). The maintenance decisions made based on this rule are shown in the last row of Table 4. As seen,

- for Liner 12 the preventive replacement will be performed at $t = 37.81$;
- for Liner 25 an inspection will be performed at $t = 54.50$ and a preventive replacement may be performed at that inspection since the inspection time is close to the actual failure time; and
- for Liner 28, an inspection will be performed at $t = 43.50$.

As a result, the decision based on the residual life distribution of an individual generally can lead to a PM action.

In the wear monitoring context, there are other decision problems, such as determining a PM threshold (see Jiang 2010), determining an opportunistic maintenance threshold (see Zhang and Zeng 2015), and dynamically determining the time of the next inspection (see Jiang 2010).

5. Conclusions

In this paper, we pointed out that the bounded assumption for an increasing degradation process is generally inappropriate. An interesting finding associated with this observation is that a model with good goodness-of-fit does not guarantee that it is a valid reliability model.

We have extended the log-linear model to a more general case. The extended log-linear model is no longer bounded, and the corresponding failure intensity can be constant, increasing or decreasing but asymptotically tends to a constant.

Though the extended log-linear model has better flexibility, it is only applicable for modeling the behavior of the first two phases in a three-phase wear process (with a bathtub-shaped wear rate). The extended model can be further modified to adapt the situation where the deterioration rate changes from a constant rate to an increasing rate (Fouladirad and Grall 2011). This is a topic for future research.

We have assumed that the degradation increment is independent but non-homogeneous. It is more reasonable to assume that the degradation increment is state-dependent for some cases. This is another topic for future research.

Acknowledgement

The research was supported by the National Natural Science Foundation of China (No. 71371035).

References

- Bocchetti D, Giorgio M, Guida M, Pulcini G (2009) A competing risk model for the reliability of cylinder liners in marine Diesel engines. *Reliability Engineering and System Safety* **94**, 1299–1307.
- Fouladirad M, Grall A (2011) Condition-based maintenance for a system subject to a non-homogeneous wear process with a wear rate transition. *Reliability Engineering and System Safety* **96**, 611–618.
- Guida M, Poli P, Pulcini G (2014) Reliability analysis and prediction of the failure process of some traction converters. *WIT Transactions on the Built Environment*, **135**, 279-290.
- Jiang R (2010) Optimization of alarm threshold and sequential inspection scheme. *Reliability Engineering and System Safety* **95(3)**, 208–215.
- Jiang R (2013) A tradeoff Bx life and its applications. *Reliability Engineering and System Safety* **113**, 1–6.
- Zhang X, Zeng J (2015) A general modeling method for opportunistic maintenance modeling of multi-unit systems. *Reliability Engineering and System Safety*, **140**, 176-190.

Improved preventive maintenance of a rotating machine by fault monitoring of its components via a neural network model

S. Khamel* and Kh. Bouacha**

* Department of Industrial Maintenance, Mohamed Chérif Messaadia University, P.O. Box 1553, Souk Ahras 41000, Algeria. khamelsamir@yahoo.fr

** Department of Industrial Maintenance, Mohamed Chérif Messaadia University, P.O. Box 1553, Souk Ahras 41000, Algeria. khbouacha@yahoo.fr

Abstract. The present work concerns a study of defects that can act during the turning operation of a machine. A mechanism simulating a gearbox has been imagined and designed in Department of Mechanical Engineering, Mohamed Chérif Messaadia University, Souk Ahras, Algeria by Dr. Khaider Bouacha. This device has permitted us to realize the experiments founded on the measurement of vibration signals corresponding to faults created on the gear and on the ball bearing (inner ring / outside ring). A fault monitoring system based on a neural network model with optimized architecture has been developed; it is powered by six scalar indicators from each of the eight segments of each signal: kurtosis, root mean square (RMS), energy, power, peak value and factor peak. The signals were taken in both radial and tangential directions. This segmentation allows the system on the one hand to have more precision in the defect detection, and secondly, to be able to locate it. Finally, the recorded detections show reliable predictive maintenance of the machine and the proposed neural network model has proven its robustness. Unfortunately, these results are specific to the off-line state and need to be tested on-line.

1. Introduction

Monitoring and diagnosis of faults are of very important in most industrial sectors. They consist of measuring physical signals on machines, industrial installations ... their analysis delivers a diagnosis on their mechanical state. This approach ensures the safety of equipment and people, the respect of the environment and the optimization of condition maintenance. The use of acoustic and vibration signals is important because it allows to understand the dynamic phenomena that occur in the operating systems (rotating machines, structures, ...). Thus It is possible to characterize most defects and malfunctions from knowledge or physical modeling of these phenomena. This step helps to bring out the signals, the "symptoms" that are highlighted by signal processing methods.

2. Fault diagnosis methods

The signal analysis methods proposed for fault diagnosis are classified into two categories: conventional methods and advanced methods. Firstly, the statistical indicators consist in analyzing time characteristics of the recorded signal, any occurrence of a fault leads to significant changes of these characteristics. They are simple, easy to interpret and can be supplemented by thresholds which triggers an alarm when exceeded, so they are adapted well for monitoring and on-line control.

Fourier analysis is one of the conventional methods; it is to decompose the energy of the signal analyzed by frequency bands, it has remarkable properties that make it an essential tool to achieve many treatments: filtering, noise removal, demodulation, etc. Thus, any current measurement is compared to a frequency template fixed to a healthy state. Fourier analysis is particularly suitable for the monitoring and diagnosis of rotating machines whose signals distribute their energy on harmonics well localized in frequency. The magnitude and position of these harmonics are a real "mechanical signature" of the machine status. Any abnormal deviation indicates a fault, Braun (1986).

The cepstral analysis is widely used for monitoring gear that exhibits a rich harmonic spectrum, Randall (1975). The other property of the cepstrum is to transform a convolution product into an addition, it can sometimes be used to separate the effect of excitation and the effect of the transfer function (deconvolution). This property, exploited in not destructive no destructive control, can be traced back to the source of the defects, Capdessus and Sidahmed (1992).

The envelope analysis allows reunite the concepts of time and frequency. It is relatively easy to implement and is much used for the monitoring and diagnosis of rotating machines (non-stationary signals) as well as non-destructive testing (acoustic emission). In some cases, the analysis of envelope is followed by a Fourier analysis (spectrum envelope) which reveals such periodic structures of the envelope. The envelope is quite recommended for the detection and identification of abnormal and repetitive shocks, Mc Fadden (1986).

The time-frequency analysis allows to characterize, in the frequency domain, non-stationary signals such as acoustic and vibration signals generated by a diesel engine during one cycle (compression, injection, combustion), Chiollaz and Favre (1990). Several methods have been developed to improve the performance of "spectrogram" which sees its "resolution" frequency decrease when one focuses on a short temporal region Courbebaisse (1995).

Among the advanced methods, we note cyclostationarity; when signals exhibit periodicities at the auto-correlation function the cyclostationarity is of 2nd order. This property allows to generate a new function called spectral correlation defined as the Fourier transform in two dimensions according to time and the cycle of the correlation function. This theory knows many developments and applications for acoustic and vibration diagnostics of machines, Capdessus and Sidahmed Lacoume (2000).

For signals that are not Gaussian but called having "Gaussian" distribution, it is necessary to use higher-order moments (3, 4, ...). This generated methods based on High Order Statistics, Raad, Antoni and Sidahmed (2003). Thus, higher order statistics allows to study the nonlinear behavior of a system. The use of higher-order statistics is often efficient for the detection, characterization and classification of defects: shock, involving non-linearity or structural type, Antoni, Daniere and Guillet (2002).

Diagnostics from the signal can be presented as inverse problems, which at the observation of an effect, seek its cause. Solving inverse problems is fundamentally difficult because of the many of unknown factors that characterize and instability of their solutions, Mansour, Barros and Ohnishi (2000). Many efforts have recently been deployed in this area, with promising success. We distinguish three sub-problems: deconvolution, blind identification and source separation, Gelle, Colas and Delaunay (2000). The deconvolution seeks to find, from the response of a system the excitation at its entrance. Original solutions have recently been made to take into account noisy measurements by the Bayesian approach and higher order statistics. The Bayesian approach can convert any breach on the direct transfer in terms of probabilities, and so stabilize the inversion system. The higher order statistics use the maximum of information contained in the various statistical signal times to increase the number of available equations. This allows certain conditions to separate the effects of the transfer of excitation and noise, Lee and NandiK (1998). The blind identification consists of estimating parameters that describe a transfer to a physical environment from external measures only, without knowing the excitations. The problem differs from the deconvolution in the sense that it is not the source that is looked for to rebuild but the function that characterizes his transfer to the sensor. This is justified if the defect corresponds to a structural deterioration (crack, inclusion, etc.) rather than exciting force. The approaches are highly dependent on application domain, Antoni, Daniere, Guillet and El Badaoui (2002). Methods of separation of sources have recently been developed. Only the statistical independence of sources is supposed, which would separate the contribution of each source in the measured signals, Gelle, Colas and Delaunay (2000). However, it is undeniable that the separation of sources potentially numerous applications in diagnosis. Source separation has been used in some applications for several separate defects that may appear simultaneously. The approaches are mainly based on higher order statistics on the information theory, but also on time-frequency treatments.

3. Experimental protocol

The experiments where realized using a device simulating a gearbox, it is composed of two shafts sufficiently rigid, fixed at the extremities by rolling bearings. They are easily demontable and are equipped with tapped hole for fixing the accelerometers. The rotation is ensured by an electric motor developing a maximum power of 0.6 KW (Figure1.).



Figure 1. Device used for the experiments

A coupling placed between the drive shaft of the device and the milling spindle, this is necessary for to isolating the structure from misalignment of the motor. The stand is also equipped with a set of identical gears and one for pinions, one of the gears is keep at the healthy state and the others are subject to different desired defects.

Experiments are based on the measurement of vibration signals collected by accelerometers absolutely fixed on the bearings; magnetic fixing was avoided to ensure accurate data acquisition.

4. Planning experiments

The use of the complete experimental plan allows the determination of the influence of the operating parameters on the vibration signals. Indeed, the principle is to vary simultaneously the levels of one or more factors in each experience. The chosen parameters and allocation of respective levels are shown in Table 1.

Level	State	Speed (rpm)	Frequency (Hz)
1	Healthy	150	800
2	One localized defect	250	1600
3	Two combined defects	350	3200
4	Three combined defects	/	6400

Table 1. Input parameters and they levels

5. Development of a neural network model

5.1 Objective of the approach

The objective of this part of this research is to propose and implement an approach for the development of a fault monitoring and diagnostic system based on a neural network model with an optimized architecture for: the number of hidden layers, the number of neurons in each hidden layer, the activation functions and the types of training algorithm. These factors are used as inputs to the optimization approach. The correlation coefficients and the mean squared error (MSE) are exploited as outputs.

5.2. Type of the proposed ANN model

The neural network model adopted in this process of development of a monitoring system, is the perceptron multilayers with a Feed-forward algorithm. Those neural networks are considered as nonlinear static systems, where each neuron in a layer is connected to all the neurons of the previous layer and the next layer

(except for layers input and output) and there are no connections between neurons of the same layer. The information propagates from layer to layer without the return back, (Figure 2).

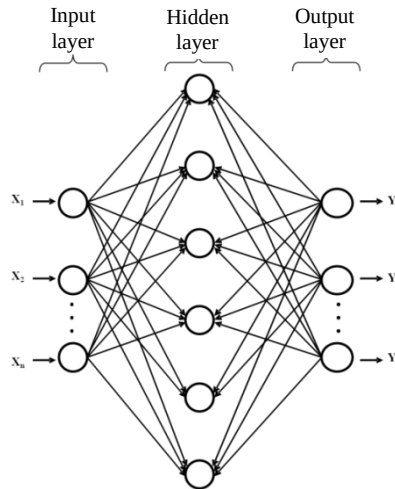


Figure 2. Configuration of ANN multi-layer (Feed-forward).

The effectiveness of this model is represented by its ability to predict the nonlinear behavior of the synthesized values and its convergence speed. Indeed, the use of neural networks for function approximation can be justified as follows:

- Implementation simplicity (a few preliminary mathematical analysis).
- Universal approximation ability proven.
- Possibility to take the process as "black box".
- Robustness relative to internal network failures.
- Ability to adapt to the conditions imposed by any environment.
- Easy to change its parameters (weight, number of neurons in the hidden layers, number of hidden layers ...) in a possible adjustment in this environment.

5.3. Neural network Inputs and outputs

The inputs used to serve the neural network are temporal indicators directly related to the 8 segments of the signals: energy, kurtosis, power, peak value, crest factor and the root mean square value (RMS). The desired outputs correspond to the binary matrix of defects detection (code : 0 or 1). The codification is presented in Table 2.

State	Defect	Codification			
Healthy state	/	1	0	0	0
Localized defect	bearing outer race	0	1	0	0
Two combined defects	bearing outer race + gear	0	1	1	0
Three combined defects	bearing outer race + gear + bearing inner race	0	1	1	1

Table 2. Defects codification (Neural network Outputs)

5.4. Training of the Neural network

All neural network models require a training. Several types of training can be adapted to a single neural network. The selection criteria are often the convergence speed and generalization performance. The training

method reserved for the present work is that by packet: the back-propagation of the error gradient algorithm associated with the Levenberg-Marquardt algorithm. The optimal architecture is the result of optimization achieved using the response surface methodology associated with the composed desirability.

- Configuration : "48-9-192".
- One hidden layer.
- 9 neurons in the hidden layer.
- Combination of activation functions: 'tansig'/'tansig'/'Purelin'.

According to the curve in figure 3, we can see that the training of the neural network model converge after only 43 iterations, a total training time not exceeding a few seconds, very short compared to other methods. The performance of the model are expressed as follows:

- Training $R_A = 0.999$.
- Validation $R_V = 0.9558$.
- Global $R_G = 0.98814$.
- Error MSE = 0.0000000567.

The value of the global correlation coefficient for the neural network model is 0.98814, this admits a relatively high significance of the model and a very good fit to the experimental data.

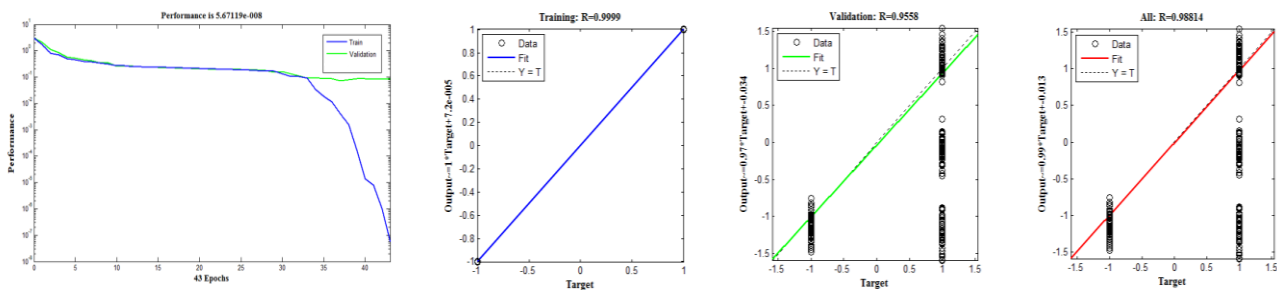


Figure 3. Performance of the neural network

6. Results analysis

The results show that the neural network model used for fault monitoring of the gearbox studied exhibits excellent robustness. This is justified by its fault detection success rate, also by the quality of information derived from the results found. Indeed, the model can recognize a localized or combined fault (such as the outer or inner ring of the bearing or the gear). The application of neural network model for fault diagnosis has given good results with remarkable accuracy. This is the fruit of the consideration of six indicators simultaneously which led to 97% detection success rate.

7. Conclusion

The results obtained during this approach for the development of a neural network model are very satisfactory. The overall success rate of monitoring and diagnostics system developed with the optimal configuration is in the order of 97% for determining the state of the machine. However this success rate is specific to the "off-line" state. Testing in real time this neural network model in a machine monitoring process is the most important of our prospects.

Références

- [1] Braun S (1986) Mechanical signature analysis. Academic Press.
- [2] Randall B (1975) Gearbox fault diagnosis using cepstrum analysis. *Proc. of the fourth world congress on machinery monitoring*, Newcastle, V1.
- [3] Capdessus C, Sidahmed M (1992) Analyse des vibrations d'engrenages: cepstre, corrélation, spectre. *Revue Traitement du Signal*, V8 (5).
- [4] Mc Fadden PD (1986) Detecting fatigue cracks in gears using amplitude and phase demodulation of meshing vibrations. *ASME J on vibrations, acoustic, stress*, V108.

- [5] Chiollaz M, Favre B (1990) Caractérisation fine de bruit moteur par analyse temps-fréquence. *Revue traitement du signal*, V8 (5).
- [6] Courbebaisse G (1995) Caractérisation d'un système d'injection par analyse temps-fréquence, *Revue Traitement du signal*, V12 (5).
- [7] Capdessus C, Sidahmed M, Lacoume JL (2000) Cyclostationary processes: applications to gear early fault diagnosis. *Mechanical Systems and Signal processing*, V14 (3).
- [8] Raad A, Antoni J, Sidahmed M (2003) Indicators of cyclostationarity: proposal, evaluation and application, *Proc ICASSP*, Honk Kong.
- [9] Antoni J, Daniere J, Guillet F (2002) Effective vibration analysis of IC engines using cyclostationarity: Part I-A methodology for condition monitoring. *Journal of Sound and Vibration*.
- [10] Mansour A, Barros AK, Ohnishi N (2000) Blind separation of sources: methods, assumptions, and applications. *IECS, trans fundamentals*, VE 83-3(8).
- [11] Gelle G, Colas M, Delaunay G (2000) Blind separation applied to rotating machines using acoustical and vibratory analysis. *MSSP*, V14 (3).
- [12] Lee JL, Nandi K (1998) Blind deconvolution of impacting signals using HOS. *MSSP Journal*, V12 (2).
- [13] Antoni J, Daniere J, Guillet F, El Badaoui M (2002) Blind separation and identification of cyclostationary processes. *Proc ICASSP*, Orlando.

Modelling Asset Management of Railway Overhead Line Equipment

P. Kilsby*, R. Remenyte-Prescott** and J.D. Andrews***

Centre for Risk and Reliability Engineering, University of Nottingham, University Park, Nottingham, NG7 2RD

*evxpk4@nottingham.ac.uk **r.remenyte-prescott@nottingham.ac.uk ***john.andrews@nottingham.ac.uk

Abstract. The overhead line equipment (OLE) is a critical sub-system of the 25kV AC overhead railway electrification system. There is a need to evaluate OLE asset management strategies through a whole life cost analysis that considers degradation processes and maintenance activities of the OLE components so that the investment required to deliver the level of performance desired by railway customers and regulators can be based on evidence from modelling results. A Petri Net model is proposed to simulate the degradation, failure, inspection and maintenance of the main OLE components and to calculate various statistics associated with the cost and reliability of the system over its whole life. The Petri Net considers all the main OLE components in one model and can simulate both fixed interval and risk based maintenance regimes. To allow such processes to be modelled accurately and efficiently, High Level Petri Net features are used. The model developed is the first of its kind, in such detail, for OLE and the applicability of Petri Nets for modelling many processes on a large system, containing numerous components, is shown.

1. Introduction

Individual OLE component failures often result in system failure and delays of the timetabled train service. Therefore, it is imperative that inspection and maintenance of the OLE takes place in order to uphold system reliability. Network Rail (the British railway infrastructure provider) aim to maintain their current assets and correctly specify the assets to be installed in new systems, so that the required outputs, such as system reliability and line speed, can be achieved at the lowest whole life cost for the system (Skinner, et al., 2011). The whole life cost of an asset is composed of its acquisition costs, associated with its design and installation, and its ownership costs, associated with its failure and maintenance (British Standards, 1997).

The OLE is a sub-system of the 25kV AC overhead electrification system, which is Network Rail's preferred electrification system, representing 63% of the electrified network which contains over 5000km of railway. With such a large electrified network and many electrification schemes planned in the near future, substantial economic savings can be realised through specifying the OLE installation types and maintenance regimes that achieve the required outputs at the lowest whole life cost. Through modelling OLE component degradation and failure over the whole life of the system, whilst also taking into account the inspection and maintenance regimes, the expected whole life cost for a given OLE installation type and maintenance regime can be estimated. The results for different project options and maintenance regimes can then be compared.

Relatively little research into asset management and whole life cost modelling of OLE has been completed to date. Chen et al. (2007) used Fault Tree Analysis to evaluate the overall reliability of a section of electrified line (including OLE components and other assets in the power system). Ho et al. (2005) described a stochastic lifetime model for estimating failure and maintenance costs for a railway traction power system. A Monte Carlo simulation of the distributions describing the component lifetimes was undertaken taking into account fixed interval preventative maintenance and corrective maintenance after a failure. Min et al. (2009) employed a similar methodology to Ho et al. (2005), but during each preventative maintenance visit, different maintenance actions were considered (such as repair, replace, do nothing) that improve the component reliability by different amounts. A genetic algorithm was then used to find the optimum maintenance strategies with respect to the overall cost and system reliability. Such models can be used to estimate the reliability and cost of OLE components and evaluate the suitability of the preventative maintenance strategy. However, the methods used are relatively simplistic and therefore do not necessarily model all the processes that can be accounted for. Furthermore, instead of fixed interval preventative maintenance visits, Network Rail implement a risk-based maintenance strategy where OLE component maintenance is scheduled based on the condition it is found to be in during routine inspections. A more sophisticated methodology is required to model such asset management strategies accurately, so that all the main processes that influence the component and whole life cost of the system are accounted for.

This paper describes a Petri Net (PN) model that has been developed to simulate the degradation, failure, inspection and maintenance of OLE components. PNs are used to represent the behaviour of a system by modelling the concurrent or asynchronous events that occur. Since Carl Petri first developed PNs in 1962, due to their power and flexibility, PNs have been used to model many processes– including railway track asset management (Andrews, 2012). British Standards (2012) have a published standard for PNs where a detailed description of the approach can be found. Like Markov models, PNs are often used to graphically represent a state-based model with the different states (known as places) referring to different conditions of the system modelled. In addition, PNs can model more complex processes (especially if the framework is extended to include High Level PN features) and the transitions between places are not limited to occur at a constant rate and obey the Markov property. The flexibility of PNs makes them an ideal choice for accurately modelling the main processes that influence the operational cost and reliability of the OLE over the whole life of the system.

2. System Description

Live conducting wires, insulators and supporting equipment installed along a railway line are collectively referred to as OLE. The main OLE components are shown below in Figure 1. Structures support all OLE components by raising them above the track. Registration equipment is attached to the structures via insulators that separate the live components from Earth. The contact and catenary wires are attached to, and aligned by, the registration equipment and the contact wire is suspended below the catenary wire by droppers. The train's pantograph rubs against the contact wire to obtain the traction power. These components exist at each structure and repeat along the entirety of an electrified line. Since the span between structures is typically about 50m, the total number of OLE components on a line is considerably large.



Figure 1. The Main OLE Components

Network Rail undertake a risk-based maintenance strategy for the OLE whereby an individual component maintenance is scheduled based on the condition it was found to be in during routine inspections. Network Rail have developed standards that describe the maximum time that maintenance must be completed by for each component defect or degradation to reduce the risk of failure. However, since maintenance engineers are responsible for many OLE components and obtaining access to and isolating the line is often logistically constrained, the components requiring maintenance in the same area are scheduled to be maintained at the same time. The size of such an area is dependent on the access to the line. In this study it is assumed that maintenance of components in an area of approximately 2 miles in length can be undertaken during the same visit. The main inspection regimes are low level walking inspections and high level intrusive inspections (assumed to be undertaken every 28 days and every 2 years respectively for high speed lines). Some component defects, such as seized fittings on the registration equipment, may not be easily identifiable during a walking inspection, and therefore they are revealed during the high level inspections only. Service affecting failures are revealed sooner, by the power supply tripping or by alerts from rails users, such as train drivers.

3. Development of a Petri Net model for OLE

The PN models the degradation, failure, maintenance and inspection processes of all of the main OLE component types (shown in Figure 1) that have a significant influence on the whole life cost of the system.

3.1 Basics of the Petri Net Model

The PN contains places that describe the conditions of the different OLE components studied and the transitions move tokens between these places to model how the components can degrade and fail and be inspected and maintained. The PN is stochastically timed and a Monte Carlo simulation is used to evaluate it. Stochastically timed transitions exist such that, after becoming enabled, they fire after a time that is based on probabilistic distributions that describe the times to reach states of failure or degradation for the component. Such distributions can be found by analysing real data. However, due to lack of high quality data, exponential and Weibull distributions have been assumed. For illustrative purposes in this paper, the firing times for inspection and maintenance transitions are fixed and based on times stated by Network Rail's maintenance standards.

3.2 A Whole System Analysis

The OLE system consists of many different components that the maintenance regimes often address at the same time. Therefore, a holistic approach is used whereby all the main components that influence system reliability are considered in the same model so that the interactions between the different components can be modelled. The fact that the degradation of one component may influence the degradation and failure rates of other components is not considered in this study, since such relationships are difficult to determine and quantify accurately with limited data. Instead, the interactions between different components are modelled by considering common intervention actions. This is primarily done by scheduling the maintenance of different components to occur at the same time, if the components are known to require maintenance. Opportunistic inspection of nearby components after failures and during maintenance actions is also modelled. The PN developed is hierarchical, with a different sub-net describing the degradation, failure, inspection and maintenance processes for each of the OLE component types studied. The different sub-nets are then linked together through the interactions, described above, for the whole system analysis.

3.3 High Level Petri Net Features

The PN developed is a High Level PN, with features that add further functionality to standard PNs to enable the processes in the system to be modelled more accurately and efficiently. Multiple different token types that can exist in the same place are used to represent different instances of each component type in the area studied. The tokens can also contain information regarding their location and scheduled maintenance times. The transitions then have different modes, so that the different token types can be modelled concurrently. A function can exist within a transition arc in order to fire a specific number of tokens to a place according to this function. Further information about the implementation of PNs containing different token types and transition modes can be found in documentation for High Level PNs (British Standards, 2010). This study used reset arcs, which make the marking of the token type in their connected place equal to a specific value, regardless of the initial marking. When tokens may exist in multiple places at the same time, reset arcs are useful for clearing them from certain places without the need for multiple transitions, so that the PN can be more concise and efficient.

3.4 Example Sub-Net for an Insulator

For each component, there exists a good or working condition of the component (denoted by P0 in Figure 2). In addition, there exists a condition where the component has degraded to the extent that, according to Network Rail's maintenance standards, maintenance should be undertaken (P1). For example, for insulators, the standards state that maintenance should be undertaken when signs of minor damage are present, i.e. when less than 3 sheds are damaged. Similarly, the component condition can deteriorate further to be severely degraded (P2). Finally, a component can also be in a failed condition (P3), which causes an interruption of the timetabled train service and delay costs are incurred. There are also places that represent the revealed degraded and revealed severely degraded condition of the components (P4 and P5 respectively), since changes in component condition are not known until after an inspection has been undertaken. A number of places are used to collect statistics, such as the number of maintenance actions in a year (P6) and the duration of time a component stays in a degraded condition (P7). P60 represents maintenance as being undertaken in

an area and is linked to every sub-net. P61 records statistics associated with the number of maintenance visits and isolation costs for the maintenance area.

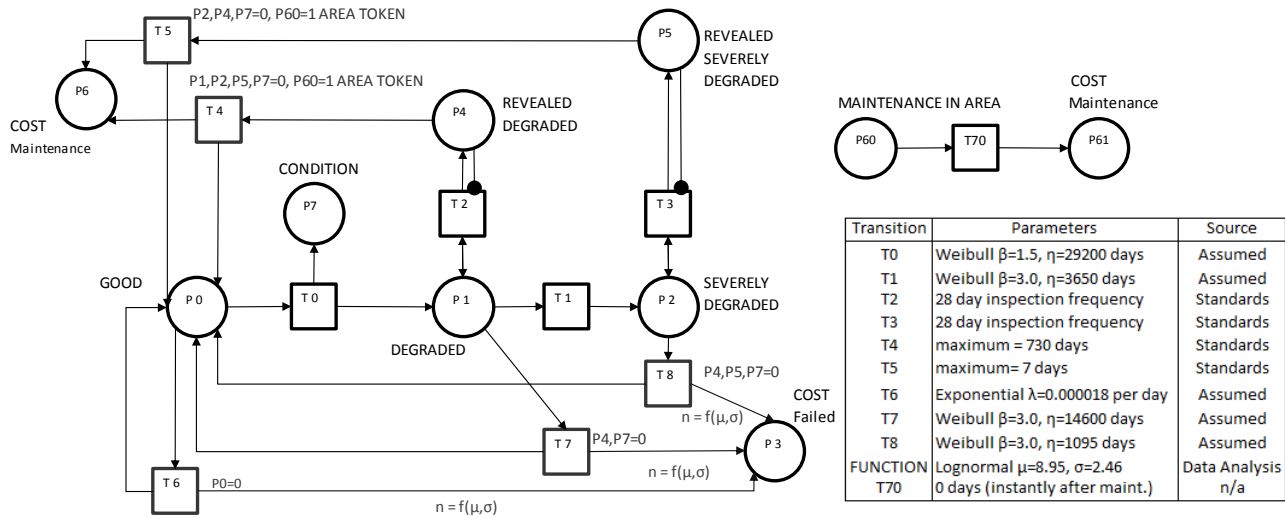


Figure 2. Insulator Sub-Net (Transitions with reset arcs list the places that are reset next to them)

Degradation transitions (T0 and T1) describe the process of degradation from the good state to reach the degraded and severely degraded states. Inspection transitions, T2 and T3, describe an inspection where the two states of degradation are revealed, at which point maintenance is scheduled for the component.

The firing time of a maintenance transition (T4 and T5) is calculated by considering whether any other components (of any type that can be grouped for maintenance) have been scheduled for maintenance in the same area. Note that rules are implemented in the code that runs the model and are not depicted in the example PN. If other maintenance is due to occur before the maximum delay allowed for the component defect type (for an insulator, 2 years if degraded, and 7 days if severely degraded), then the maintenance of the component is undertaken earlier, i.e. with other planned maintenance in the area. Otherwise, the maintenance is planned according to the maximum time in the standards. The date that this maintenance is due to occur is recorded, so if other components are found to require maintenance, they are scheduled to be maintained at the same time. Maintenance returns components to the good condition (P0), for insulators a like for like replacement is undertaken. The maintenance transition reset arcs clear all other instances of the relevant token type from any of the other places it could have moved to (due to further degradation and inspection). A reset arc also sets the marking of the corresponding area token type (that represents the area that the maintained component is in) in P60 to be equal to 1, to signify that maintenance is underway in the area. T70 will fire, once all the maintenance transitions that were scheduled to occur in the same area have fired, and statistics related to the number of maintenance visits and isolation costs for the area are in P61.

There are 3 failure transitions (T6, T7 and T8) that represent component failure from each of the component conditions that are modelled. When the component is in a good condition, failure will occur primarily due to random processes, such as bird strikes. When the component has degraded, there is an increased chance of failure due to its degradation. Due to lack of available data, downtime is not considered in the analysis. Instead, the cost of failure is used to account for different failure severities since data containing the delay costs due to failures of the different component types was available. A distribution of failure costs was found for each component type (the lognormal distribution provided an appropriate fit). When a failure transition fires, the cost of the failure is sent to P3, which records the failure statistics. This is done using an arc with a function that samples the failure cost from the lognormal distribution of failure costs for an insulator, and fires the cost to P3. It is assumed that all the different failures for an insulator are revealed and have the same distribution of failure costs. The failure transitions fire the cost to P3 and also fire a token back to the good condition as the component is replaced after failure.

The distributions of time for degradation (T0 and T1) and failure (T6, T7 and T8) can be found by performing survival analysis of historical maintenance and failure data, as carried out by Meier-Hirmer et al. (2006) for the OLE in France. Due to limitations in the data, some of the parameter values for Weibull and exponential distributions are assumed in this paper, as shown in Figure 2. Nevertheless, Weibull distribution

represents wear-out failures in this study, since an increasing hazard rate was observed in the data and the characteristic life was estimated using the data. Exponential distribution represents random failures, such as bird strikes, and a constant hazard rate is assumed for T6.

3.5 The OLE System Petri Net

A sub-net has been developed for each component type considered in this study. Each sub-net contains a different token type for each instance of the component in the area studied, and each transition has a different mode for each token type. Figure 3 shows an overview of how the different sub-nets interact in the system PN. The maintenance of components in the same area is scheduled to occur at the same time (apart from maintenance of structures). When maintenance is undertaken, the line needs to be isolated and the work undertaken by the maintenance team. Therefore, all the maintenance transitions fire a token to a sub-net that models this process (P60, T70 and P61 in Figure 2). Additionally, when maintenance of a component takes place, the adjacent component degradation will be revealed by a detailed inspection, therefore component maintenance transitions also fire a token with location information to a sub-net that models this process immediately after the maintenance transition has fired. Similarly, after the component fails, a low level inspection of the nearby section of line (roughly 1 mile) takes place immediately afterwards.

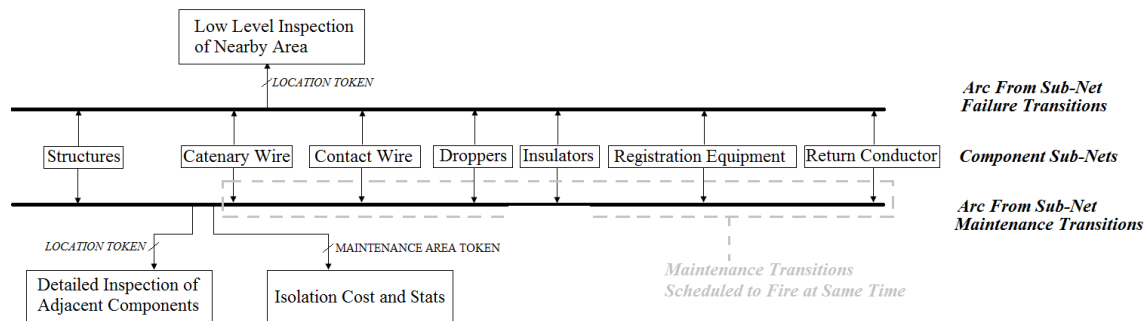


Figure 3. Overview of System Petri Net

4. Example Outputs of the Model

The model was used to simulate 100 years of operation of all the main OLE components in a maintenance area, which is approximately 2 miles long and contains 53 structures. The 100 year timeframe is chosen to allow the entire life cycle of most of the components to be witnessed (as OLE component lifetimes are typically between 40 and 80 years). The system PN developed contained 64 places, 69 transitions and 384 tokens (resulting in equivalent to 3746 transitions at runtime due to different modes for each transition). Using C++, bespoke software has been developed to perform the Monte Carlo simulation of the PN and collect various statistics. Each simulation was coded to be performed in parallel in order to reduce the run time. 100,000 simulations of the model were completed in less than 1 minute using a computer with a dual core Intel i3 processor. The mean yearly total maintenance and failure cost had converged in this number of simulations.

A number of output statistics, that express the system behaviour and various costs incurred over the 100 year period, were obtained. The statistics were found for each component type and for the overall OLE system in the maintenance area studied (by collating the different results). For example, in Figure 4, the left-hand chart shows the mean yearly maintenance costs of all the insulators in the area. The Monte Carlo simulation allows statistics, expressing the large degree of uncertainty associated with the costs, to be calculated, such as the inter quartile range of the total cumulative cost (comprised of inspection, maintenance and failure costs), as shown in the central chart. The right-hand chart shows that yearly failure costs are greater than maintenance costs for the system. The yearly maintenance and failure costs and failure rate increase overtime, as components age and degrade (and large scale renewals were not considered in this study). Note that the fluctuations between alternate years are due to the scheduling of the high-level inspections (modelled to occur every 2 years). Therefore, component degradation is only revealed every 2 years, and maintenance and component failures are more likely to occur every other year. With many of the initial failures being due to random failure or components with degradation that is revealed through a low

level inspection (such as insulators), the failure rate does not fluctuate noticeably until after 40 years (see right-hand chart). After 40 years the components with degradation revealed through high level inspections (such as the catenary wire) are more likely to be in a degraded state and therefore more likely to fail.

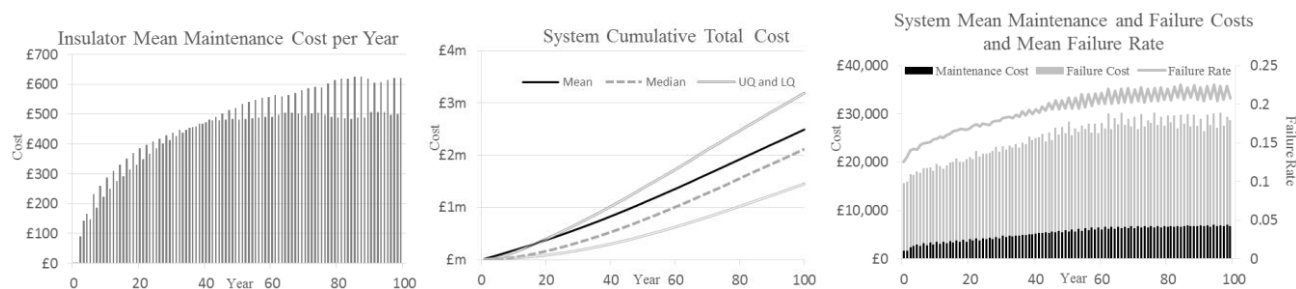


Figure 4. Charts Plotting Some of the Statistics Calculated.

5. Conclusion

This paper demonstrates how a PN approach can be used to evaluate the OLE asset management strategies through a whole life cost analysis. Degradation and failure processes are modelled for all the main OLE component types, whilst also taking into account the inspection and maintenance actions undertaken as part of a risk based maintenance regime. The whole system model considers a large number of components in a maintenance area and allows opportunistic maintenance and inspection to be modelled. The variety of statistics calculated can be used to support decisions on future maintenance requirements and to evaluate the component and system reliability over the whole life of the system. Further work could consider the influence different component reliabilities have on system reliability and whole life cost using a sensitivity analysis of the component degradation and failure rates. The PN could also be run in conjunction with a genetic algorithm to find the optimum inspection frequencies and maintenance strategies.

Acknowledgements

The project is supported by Network Rail and the Engineering and Physical Sciences Research Council (EPSRC). The authors gratefully acknowledge the support of these organisations.

References

- Andrews, J.D. (2013). A modelling approach to railway track asset management. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 227, 56-73.
- British Standards (1997). BS 5760: part 23, Guide to Life Cycle Costing.
- British Standards (2010). BS ISO/IEC 15909-1:2004+A1:2010, Systems and software engineering- High-level Petri nets.
- British Standards (2012). BS EN 62551:2012, Analysis techniques for dependability- Petri net techniques.
- Chen, S. K., Ho, T. K., & Mao, B. H. (2007). Reliability evaluations of railway power supplies by fault-tree analysis. *Electric Power Applications, IET*, 1(2), 161-172.
- Ho, T. K., Chi, Y. L., Ferreira, L., Leung, K. K., & Siu, L. K. (2006). Evaluation of maintenance schedules on railway traction power systems. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 220(2), 91-102.
- Meier-Hirmer, C., Sourget, F., & Roussignol, M. (2006). Lifetime estimation of catenary components. In *Proceedings of the European Safety and Reliability Conference, ESREL 2006*, Estoril, Portugal, Safety and reliability for managing risk – Soares, C.G. and Zio, E. (eds); 2006 Taylor and Francis Group, London, 929-934
- Min, L. X., Yong, W. J., Yuan, Y., & Yan, X. W. (2009). Multiobjective optimization of preventive maintenance schedule on traction power system in high-speed railway. In *Reliability and Maintainability Symposium, 2009. RAMS 2009. Annual*, 365-370.
- Petri, C. A. (1962). Kommunikation mit automaten. PhD Thesis, University of Bonn.
- Skinner, M., Kirwan, A., & Williams, J. (2011). Challenges of developing whole life cycle cost models for Network Rail's top 30 assets. In *IET and IAM Asset Management Conference 2011*.

Condition-based maintenance at both scheduled and unscheduled opportunities

Sz. Kálosi^{*†}, S. Kapodistria^{*‡}, J.A.C. Resing^{*§}

^{*} Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands

[†] email address: s.kalosi@tue.nl, [‡] email address: s.kapodistria@tue.nl, [§] email address: j.a.c.resing@tue.nl

Abstract Motivated by original equipment manufacturer (OEM) service and maintenance practices we consider a single component subject to replacements at failure instances and two types of preventive maintenance opportunities: scheduled, which occur due to periodic system reviews of the equipment, and unscheduled, which occur due to failures of other components in the system. Modelling the state of the component appropriately and incorporating a realistic cost structure for corrective maintenance as well as condition-based maintenance (CBM), we derive the optimal CBM policy. In particular, we show that the optimal long-run average cost policy for the model at hand is a control-limit policy, where the control limit depends on the time until the next scheduled opportunity. Furthermore, we explicitly calculate the long-run average cost for any given control-limit time dependent policy and compare various policies numerically.

1. Introduction

When planning CBM strategies, see, e.g., Jardine et al. (2006), Jardine and Tsang (2013), Prajapati et al. (2012), the typical assumption in the literature is that the asset at hand is monitored continuously and one can intervene and repair the asset at any given moment. However, in practice, especially for highly complicated high-tech systems, such as off-shore wind turbines, baggage handling systems in airports and interventional X-ray machines, downtimes are extremely costly and scheduling maintenance tasks is challenging, see, e.g., Kobbacy and Murthy (2008), Manzini et al. (2010). One solution suggested in the literature to overcome this problem is the planning of opportunistic maintenance. Most of the works on opportunistic maintenance consider only unscheduled opportunities, rarely treating the case of scheduled opportunities, see, e.g., Wang (2002) and the references therein. Different from existing research on opportunistic maintenance policies, we consider both scheduled and unscheduled opportunities for CBM. Furthermore, we consider distinct costs for maintenance based on the type of the opportunity. For a single-component model we derive the optimal long-run average cost CBM policy and calculate the corresponding long-run average cost. Thereafter, we compare it with the corresponding cost for only one type of opportunity. Our work can be viewed as a variation of the work of Zhu et al. (2016) in the direction of CBM.

The paper is organised as follows. In Section 2, we present the model and describe the exact setting that motivated our work. In the subsequent two sections we prove the two main results: in Section 3, we derive the optimal policy and in Section 4, we calculate the long-run average cost. In Section 5, we numerically compare the optimal replacement policy to other policies. Finally, in the last section, we present some concluding remarks and ideas on future research.

2. Model formulation

We consider a single component that is monitored continuously and whose condition is fully observable. We assume that the condition of the component can only degrade over time and it can be classified as *perfect*, *satisfactory* and *failed*. This type of models are known in the literature as delay time degradation models, see (Kobbacy and Murthy 2008, Chapter 14). We will refer to the state of perfect condition as state 2, the state of satisfactory condition as state 1 and the failure state as state 0. Furthermore, we assume that as soon as a component failure occurs, the component is instantaneously replaced by an “as good as new” component. So, in the mathematical formulation of the model, we may assume, due to the instantaneous replacement at failure, that the model evolves between only states 1 and 2. The component spends an exponential amount of time with rate μ_i in state i , $i \in \{1, 2\}$.

We assume that we have two types of opportunities in which we can preventively replace the component before failure: the scheduled and the unscheduled opportunities. The scheduled opportunities correspond to pre-arranged opportunities occurring according to a fixed schedule. These opportunities can be attributed to

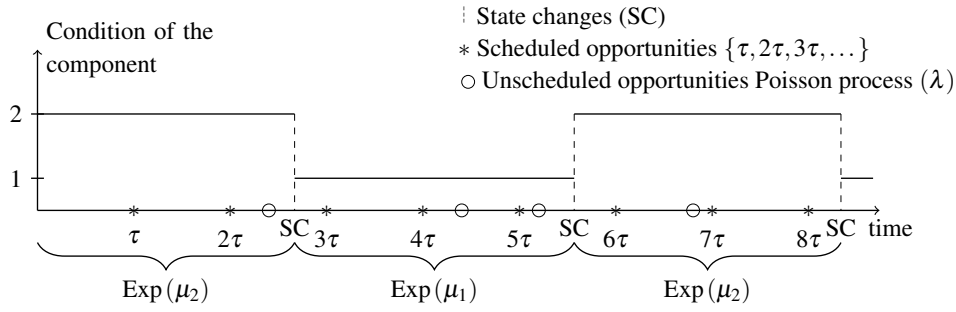


Figure 1. A sample path of the model.

either service/maintenance agreements or to regulation imposition checks. We assume that the scheduled opportunities occur at epochs $\tau, 2\tau, 3\tau, \dots$, with $\tau > 0$. This is also in accordance with what happens in practice as maintenance actions once planned are typically not rescheduled. The unscheduled opportunities correspond to random opportunities triggered by failures of other unrelated components of the same system or failures of other systems in close proximity. We assume that these unscheduled opportunities occur according to a Poisson process at rate λ .

The unscheduled and scheduled opportunities, abbreviated by USO and SO, respectively, serve as opportunities for the monitored component to be replaced preventively. Such a preventive replacement is assumed to cost less than a corrective replacement upon failure, which costs say c_c . Moreover, incorporating a planning perspective, we may assume that the preventive replacement cost at a scheduled opportunity, say c_p^{so} , is less than or equal to the corresponding cost at an unscheduled opportunity, say c_p^{uso} , that is $0 < c_p^{so} \leq c_p^{uso} < c_c$.

Our aim is to determine a policy when to replace the component based on its condition and the opportunity type, i.e. scheduled or unscheduled. More explicitly, we will need to formally define the state space, which refers to the condition of the component, the action space and the decision epochs. The state space is governed by the process depicting the condition of the component, i.e. the Markov chain evolving between the states $\{1, 2\}$. The action space consists of only two actions: replace the component by an “as good as new” or do nothing. Lastly, the decision epochs are the epochs of the scheduled and unscheduled opportunities. In Figure 1, we depict SO by (*) and USO by (o).

3. Derivation of the optimal policy

This section is devoted to the proof of our first main result on the optimal policy on when to replace the component. To this purpose, we set up our problem as a Markov Decision Problem (MDP). Due to the stochastic nature of the problem, it does not suffice to know the type of the decision epoch (scheduled or unscheduled opportunity), but it is also required to keep track of the remainder time till the next scheduled opportunity. The remaining time until the next scheduled opportunity may impact our decision, i.e. the optimal policy may depend on the residual time till the scheduled opportunity. Thus, for the full description of the condition (state) of the component, we will use a triplet descriptor $\mathcal{S} = \{(i, j, t) : i \in \{1, 2\}, j \in \{SC, SO, USO\}, t \in [0, \tau]\}$, where i indicates the condition of the component. If $j = SC$, then this means that the condition of the system is about to change and there is no decision associated with this epoch, while if $j = SO$ or $j = USO$, this means that this is a decision moment of the type of a scheduled or unscheduled opportunity. Finally, the third element indicates the remaining time until the scheduled opportunity. Note that if $j = SO$ then $t = 0$. The introduction of the time in the full description of the condition of the component renders the model inhomogeneous, and for this reason we use techniques that stem from semi-Markov Decision Problems (SMDP).

3.1. Equal preventive costs

In case of equal costs, the long-run optimal average cost policy does not depend on time. In particular, the optimal policy for state 2 is to do nothing and for state 1 the optimal policy depends on the model parameters.

Theorem 1. *Under the assumption that $c_p^{so} = c_p^{uso} = c_p > 0$, the long-run optimal average cost policy is: For state 2 to do nothing. For state 1 to replace (at both scheduled and unscheduled opportunities) if $(\mu_1 + \mu_2)c_p < \mu_1 c_c$, and to do nothing otherwise.*

For the proof of Theorems 1 and 2, we derive the policy that minimises the long-run average cost per time unit. To this purpose, we state the Bellman optimality equations for the decision process with action space $\mathcal{A} = \{\text{replace, do nothing}\}$, see, e.g., Puterman (2005), Ross (1970).

Proposition 1 (Puterman, Section 11.4.). *For a decision process with an embedded unichain, there exists a scalar g and a value function $V(\cdot)$ satisfying the system of equations*

$$V(i, j, t) = \min_{a \in \mathcal{A}} \left\{ c(i, j, t; a) - g s(i, j, t; a) + \sum_{k, l} \int_y p_{i, j, k, l}(t, y; a) V(k, l, y) dy \right\}, (i, j, t) \in \mathcal{S}, \quad (1)$$

where $s(i, j, t; a)$ is the expected sojourn time of the SMDP in state (i, j, t) under action $a \in \mathcal{A}$, and $c(i, j, t; a)$ is the stationary total expected cost of the SMDP in state (i, j, t) between two consecutive decision epochs under action $a \in \mathcal{A}$. The existence of such a scalar g ensures the existence of a policy that minimises the long-run average cost.

Proof. [Proof of Theorem 1] The Bellman optimality equations (1) for the model at hand become:

$$\begin{aligned} V(2, \text{SC}, t) = & 0 - g \int_0^t e^{-(\mu_1 + \lambda)x} dx + V(1, \text{SO}, 0) \int_t^\infty (\mu_1 + \lambda) e^{-(\mu_1 + \lambda)x} dx \\ & + \int_0^t \left(\frac{\mu_1}{\mu_1 + \lambda} V(1, \text{SC}, t - x) + \frac{\lambda}{\mu_1 + \lambda} V(1, \text{USO}, t - x) \right) (\mu_1 + \lambda) e^{-(\mu_1 + \lambda)x} dx \\ = & e^{-(\mu_1 + \lambda)t} \left(\int_0^t (\mu_1 V(1, \text{SC}, y) + \lambda V(1, \text{USO}, y) - g) e^{(\mu_1 + \lambda)y} dy + V(1, \text{SO}, 0) \right), \end{aligned} \quad (2)$$

$$V(1, \text{SC}, t) = c_c + e^{-(\mu_2 + \lambda)t} \left(\int_0^t (\mu_2 V(2, \text{SC}, y) + \lambda V(2, \text{USO}, y) - g) e^{(\mu_2 + \lambda)y} dy + V(2, \text{SO}, 0) \right), \quad (3)$$

$$\begin{aligned} V(i, \text{USO}, t) = & \min \left\{ c_p^{\text{USO}} + e^{-(\mu_2 + \lambda)t} \left(\int_0^t (\mu_2 V(2, \text{SC}, y) + \lambda V(2, \text{USO}, y) - g) e^{(\mu_2 + \lambda)y} dy + V(2, \text{SO}, 0) \right); \right. \\ & \left. e^{-(\mu_i + \lambda)t} \left(\int_0^t (\mu_i V(i, \text{SC}, y) + \lambda V(i, \text{USO}, y) - g) e^{(\mu_i + \lambda)y} dy + V(i, \text{SO}, 0) \right) \right\}, i = 1, 2, \end{aligned} \quad (4)$$

$$\begin{aligned} V(i, \text{SO}, 0) = & \min \left\{ c_p^{\text{SO}} + e^{-(\mu_2 + \lambda)\tau} \left(\int_0^\tau (\mu_2 V(2, \text{SC}, y) + \lambda V(2, \text{USO}, y) - g) e^{(\mu_2 + \lambda)y} dy + V(2, \text{SO}, 0) \right); \right. \\ & \left. e^{-(\mu_i + \lambda)\tau} \left(\int_0^\tau (\mu_i V(i, \text{SC}, y) + \lambda V(i, \text{USO}, y) - g) e^{(\mu_i + \lambda)y} dy + V(i, \text{SO}, 0) \right) \right\}, i = 1, 2. \end{aligned} \quad (5)$$

Note that in (4) and (5), inside the minimum, the left terms correspond to the action “replace”, while the right terms correspond to the action “do nothing”. Furthermore, for $i = 2$, (4) and (5) yield that it is never optimal to replace. In order to decide if it is optimal or not to replace the component in state $i = 1$ at scheduled and unscheduled opportunities, cf. (4) and (5), we define the following auxiliary functions, for $i = 1, 2$,

$$F_i(t) = e^{-(\mu_i + \lambda)t} \left(\int_0^t (\mu_i V(i, \text{SC}, y) + \lambda V(i, \text{USO}, y) - g) e^{(\mu_i + \lambda)y} dy + V(i, \text{SO}, 0) \right), t \in [0, \tau], \quad (6)$$

and rewrite (2)–(5), for $t \in [0, \tau]$,

$$V(1, \text{SC}, t) = c_c + F_2(t), \quad V(2, \text{SC}, t) = F_1(t), \quad (7)$$

$$V(i, \text{USO}, t) = \min \left\{ F_i(t), c_p^{\text{USO}} + F_2(t) \right\}, i = 1, 2, \quad (8)$$

$$V(i, \text{SO}, 0) = \min \left\{ F_i(\tau), c_p^{\text{SO}} + F_2(\tau) \right\}, i = 1, 2. \quad (9)$$

From this point onward in the proof, we will use that $c_p^{\text{USO}} = c_p^{\text{SO}} = c_p$. At a scheduled opportunity, we can either replace a component in state 1 or do nothing. We first assume that we replace the component at a scheduled opportunity, i.e. $F_1(\tau) > c_p + F_2(\tau)$, thus, Equation (9) for $i = 1$ implies that $V(1, \text{SO}, 0) = c_p +$

$F_2(\tau)$. Furthermore, for $t = 0$, Equation (6) yields $F_1(0) = V(1, \text{SO}, 0)$. Combining the last two results yields

$$F_1(0) - F_2(0) = c_p < F_1(\tau) - F_2(\tau). \quad (10)$$

Since the functions $F_i(t)$ are continuous functions in $t \in [0, \tau]$ and their difference at $t = \tau$ is greater than c_p at τ , cf. Equation (10), there exists an $\varepsilon > 0$ such that

$$F_1(t) - F_2(t) \geq c_p, \quad t \in (\tau - \varepsilon, \tau]. \quad (11)$$

Taking the derivative in (6) and after straightforward manipulations, using (7)–(9) together with (11), yields

$$\begin{aligned} F_1'(t) - F_2'(t) &= -(\mu_1 + \lambda)F_1(t) + \mu_1 V(1, \text{SC}, t) + \lambda V(1, \text{USO}, t) \\ &\quad + (\mu_2 + \lambda)F_2(t) - \mu_2 V(2, \text{SC}, t) - \lambda V(2, \text{USO}, t) \\ &= -(\mu_1 + \mu_2 + \lambda)(F_1(t) - F_2(t)) + \lambda c_p + \mu_1 c_c, \quad t \in (\tau - \varepsilon, \tau]. \end{aligned} \quad (12)$$

The solution to the above differential equation reads

$$\begin{aligned} F_1(t) - F_2(t) &= (F_1(\tau) - F_2(\tau))e^{(\mu_1 + \mu_2 + \lambda)(\tau - t)} - (\lambda c_p + \mu_1 c_c) \int_t^\tau e^{(\mu_1 + \mu_2 + \lambda)(x - t)} dx \\ &= \frac{\lambda c_p + \mu_1 c_c}{\mu_1 + \mu_2 + \lambda} + \left(F_1(\tau) - F_2(\tau) - \frac{\lambda c_p + \mu_1 c_c}{\mu_1 + \mu_2 + \lambda} \right) e^{(\mu_1 + \mu_2 + \lambda)(\tau - t)}, \quad t \in (\tau - \varepsilon, \tau]. \end{aligned} \quad (13)$$

Note that, for $t \in (\tau - \varepsilon, \tau]$ the function $F_1(t) - F_2(t)$ is monotone. Moreover, if $F_1(\tau) - F_2(\tau) - \frac{\lambda c_p + \mu_1 c_c}{\mu_1 + \mu_2 + \lambda} \geq 0$ or equivalently if $F_1(t) - F_2(t)$ is non-increasing, we can extend the above approach throughout the interval $[0, \tau]$ and show that the solution of the differential equation (13) is valid in the entire interval $[0, \tau]$, but that would contradict assumption (10). Hence, for $t \in (\tau - \varepsilon, \tau]$ the function $F_1(t) - F_2(t)$ is increasing and

$$c_p < F_1(\tau) - F_2(\tau) < \frac{\lambda c_p + \mu_1 c_c}{\mu_1 + \mu_2 + \lambda}. \quad (14)$$

Since the function $F_1(t) - F_2(t)$ is increasing for $t \in (\tau - \varepsilon, \tau]$, we now need to identify the ε , more concretely the point $t = \tau - \varepsilon \geq 0$ such that $F_1(t) - F_2(t) = c_p$. If $t = \tau - \varepsilon$ is strictly positive this contradicts assumption (10), yielding $\varepsilon = \tau$. All in all, if the policy at a scheduled opportunity is to replace the component in state 1, then the optimal long-run average cost policy at an unscheduled opportunity is to replace in state 1 if $(\mu_1 + \mu_2)c_p < \mu_1 c_c$, cf. (14).

Similarly to the case that we replace the component at a scheduled opportunity, we may assume that we do not replace the component at a scheduled opportunity and prove that then, it is also optimal not to replace the component at an unscheduled opportunity if $(\mu_1 + \mu_2)c_p > \mu_1 c_c$. At the equality $(\mu_1 + \mu_2)c_p = \mu_1 c_c$, we are cost-wise indifferent between the two actions. $\square$

3.2. Different preventive costs

In case of different preventive costs, the optimal policy, in certain cases, becomes time-dependent, i.e. there exists a threshold t^* such that it is optimal to replace if the residual time until the next scheduled opportunity is greater than or equal to the threshold and do nothing otherwise.

Theorem 2. *Under the assumption that $c_p^{\text{SO}} < c_p^{\text{USO}}$, the optimal long-run average cost policy is: For state 2 to do nothing. For state 1 to replace at scheduled opportunities and at unscheduled opportunities occurring in $[t^*, \tau)$, with $t^* = \max \left\{ 0, \frac{1}{\mu_1 + \mu_2} \ln \left(\frac{(\mu_1 + \mu_2)c_p^{\text{SO}} - \mu_1 c_c}{(\mu_1 + \mu_2)c_p^{\text{USO}} - \mu_1 c_c} \right) \right\}$, if $(\mu_1 + \mu_2)c_p^{\text{SO}} < \mu_1 c_c$, and to do nothing otherwise.*

Proof. Similarly to the proof of Theorem 1, we need to make certain assumptions here regarding the actions at the given opportunities. In particular, we distinguish three cases, each corresponding to a different set of actions. *Case (i):* $F_1(\tau) - F_2(\tau) > c_p^{\text{USO}}$; *Case (ii):* $c_p^{\text{USO}} \geq F_1(\tau) - F_2(\tau) > c_p^{\text{SO}}$; *Case (iii):* $F_1(\tau) - F_2(\tau) \leq c_p^{\text{SO}}$. The proof of this theorem is identical in structure to the proof of Theorem 1 and for this reason it is omitted. $\square$

4. Derivation of the long-run average cost

In the previous section, we derived the optimal long-run average cost policy. In this section, for any given time dependent control limit policy, we calculate the long-run average cost per time unit in Theorem 3.

Theorem 3. Consider a given policy that states that in state 2 we do nothing, and in state 1 we replace at scheduled opportunities and at unscheduled opportunities for which the remaining time until the next scheduled opportunity is greater than $\tilde{t} \in (0, \tau)$, and we do nothing otherwise. Under this given policy, the long-run average cost is

$$\begin{aligned} & \frac{1}{\tau} \left[c_c \mu_1 \left(\tilde{t} \frac{\mu_2}{\mu_1 + \mu_2} + \left(\frac{\mu_2}{\lambda + \mu_1 + \mu_2} - \frac{\mu_2}{\mu_1 + \mu_2} - \frac{\mu_2}{\lambda + \mu_1 + \mu_2} e^{(\lambda + \mu_1 + \mu_2)(\tilde{t} - \tau)} \right) \frac{1}{\mu_1 + \mu_2} \left(1 - e^{-(\mu_1 + \mu_2)\tilde{t}} \right) \right) \right. \\ & \quad + \left(c_p^{USO} \lambda + c_c \mu_1 \right) \frac{\mu_2}{\lambda + \mu_1 + \mu_2} \left(\tau - \tilde{t} - \frac{1}{\lambda + \mu_1 + \mu_2} \left(1 - e^{(\lambda + \mu_1 + \mu_2)(\tilde{t} - \tau)} \right) \right) \\ & \quad \left. + c_p^{SO} \left(\frac{\mu_2}{\mu_1 + \mu_2} + \left(\frac{\mu_2}{\lambda + \mu_1 + \mu_2} - \frac{\mu_2}{\mu_1 + \mu_2} - \frac{\mu_2}{\lambda + \mu_1 + \mu_2} e^{(\lambda + \mu_1 + \mu_2)(\tilde{t} - \tau)} \right) e^{-(\mu_1 + \mu_2)\tilde{t}} \right) \right]. \end{aligned} \quad (15)$$

Proof. For the calculation of the long-run average cost, we employ techniques from renewal-reward theory. To this purpose, we consider as renewal epochs the instants of the scheduled opportunities, thus, the inter-renewal times are deterministic and equal to τ . For the determination of the average cost during a renewal cycle, we define the following probabilities

$$p_i(t; \Lambda) = \mathbb{P}(\text{machine is in state } i \mid \text{time until next scheduled opportunity is } t), \quad (16)$$

where $t \in [0, \tau)$ and where Λ takes the values λ or 0 depending on whether or not we replace at an unscheduled opportunity given that the time until the next scheduled opportunity is t . Then, the average cost in a cycle can be obtained as follows

$$c_p^{SO} p_1(0; 0) + \left(c_p^{USO} \lambda + c_c \mu_1 \right) \int_{\tilde{t}}^{\tau} p_1(t; \lambda) dt + c_c \mu_1 \int_0^{\tilde{t}} p_1(t; 0) dt,$$

from which, once we calculate the unknown probability $p_1(t; \Lambda)$, $t \in [0, \tau)$, we can immediately derive the long-run average cost by simply dividing by τ .

The probabilities in (16) satisfy the following system of differential equations

$$p_1'(t; \Lambda) = (\Lambda + \mu_1) p_1(t; \Lambda) - \mu_2 p_2(t; \Lambda) \text{ and } p_2'(t; \Lambda) = \mu_2 p_2(t; \Lambda) - (\Lambda + \mu_1) p_1(t; \Lambda). \quad (17)$$

Solving (17), knowing that $p_1(t; \Lambda) + p_2(t; \Lambda) = 1$ for all t , $\lim_{t \rightarrow \tau^-} p_1(t; \lambda) = 0$ and $\lim_{t \rightarrow \tilde{t}^-} p_i(t; 0) = \lim_{t \rightarrow \tilde{t}^+} p_i(t; \lambda)$, $i = 1, 2$, i.e. they are continuous functions, we obtain

$$p_1(t; \Lambda) = \frac{\mu_2}{\mu_1 + \mu_2} + \left(\frac{\mu_2}{\lambda + \mu_1 + \mu_2} - \frac{\mu_2}{\mu_1 + \mu_2} - \frac{\mu_2}{\lambda + \mu_1 + \mu_2} e^{(\lambda + \mu_1 + \mu_2)(\tilde{t} - \tau + \frac{\Lambda}{\lambda}(t - \tilde{t}))} \right) e^{\frac{\lambda - \Lambda}{\lambda}(\mu_1 + \mu_2)(t - \tilde{t})}.$$

□

4.1. Special cases

In case of only *scheduled opportunities*, which corresponds to the case $\tilde{t} \rightarrow \tau$ or equivalently to the case $\lambda \rightarrow 0$, the long-run expected cost equals $\frac{1 - e^{-(\mu_1 + \mu_2)\tau}}{(\mu_1 + \mu_2)\tau} \mu_2 \left(c_p^{SO} - c_c \frac{\mu_1}{\mu_1 + \mu_2} \right) + c_c \frac{\mu_1 \mu_2}{\mu_1 + \mu_2}$.

In case of only *unscheduled opportunities*, which corresponds to the case $\tau \rightarrow \infty$, the condition of the component can be fully described using a double descriptor $\mathcal{S} = \{(i, j) : i \in \{1, 2\}, j \in \{\text{SC}, \text{USO}\}\}$ which is independent of time, and thus the new model formulation falls into the framework of regular MDPs. It can be easily shown that: For state 2 the optimal policy is to do nothing. For state 1 the optimal policy is to replace if $(\mu_1 + \mu_2) c_p^{USO} < \mu_1 c_c$ and to do nothing otherwise. Furthermore, under the optimal policy the average long-run cost is equal to $\frac{c_p^{USO} \lambda \mu_2 + c_c \mu_1 \mu_2}{\lambda + \mu_1 + \mu_2}$.

In case of only *corrective replacements*, the long-run average cost is equal to $c_c \frac{\mu_1 \mu_2}{\mu_2 + \mu_1}$.

λ	$c_p^{so} = 4000$ and $r^* \approx 1.6$									$c_p^{so} = 6500$ and $r^* \approx 1.27$									$c_p^{so} = 9000$ and $r^* \approx 0.63$								
	$\tau = 1$			$\tau = 2$			$\tau = 4$			$\tau = 1$			$\tau = 2$			$\tau = 4$			$\tau = 0.5$			$\tau = 1$			$\tau = 2$		
	a_{opt}	a_{SO}	a_{alw}	a_{opt}	a_{SO}	a_{alw}	a_{opt}	a_{SO}	a_{alw}	a_{opt}	a_{SO}	a_{alw}	a_{opt}	a_{SO}	a_{alw}	a_{opt}	a_{SO}	a_{alw}	a_{opt}	a_{SO}	a_{alw}	a_{opt}	a_{SO}	a_{alw}	a_{opt}	a_{SO}	a_{alw}
0.1	2840.41	2885.56	3384.70	3384.86	3422.03	3802.49	3807.90	3823.32	3378.56	3403.48	3719.49	3720.28	3738.77	3979.00	3985.81	3989.58	3792.57	3797.66	3916.43	3916.70	3921.39	4052.18	4055.71	4055.51			
0.5	2840.41	3042.07	3384.09	3384.86	3538.91	3784.63	3807.90	3867.21	3378.56	3489.66	3716.57	3720.28	3796.18	3956.81	3985.81	3998.72	3792.57	3816.41	3915.40	3916.70	3937.26	4039.84	4055.71	4053.46			
1	2840.41	3194.24	3383.38	3384.86	3636.35	3768.42	3807.90	3899.32	3378.56	3573.11	3713.40	3720.28	3842.96	3937.06	3985.81	4003.48	3792.57	3836.68	3914.21	3916.70	3951.98	4027.59	4055.71	4049.58			
2	2840.41	3401.88	3382.15	3384.86	3747.82	3747.68	3807.90	3932.53	3378.56	3686.18	3708.32	3720.28	3894.71	3912.27	3985.81	4006.06	3792.57	3868.78	3912.11	3916.70	3970.48	4010.20	4055.71	4041.61			

Table 1: Long-run average cost varying λ , τ and c_p^{so} , while keeping all other parameters fixed for three policies.

5. Numerical results

As a small numerical example, due to shortage of space, we examine the effect of τ , λ and c_p^{so} on the long-run average cost, while keeping all other parameters fixed. We choose $c_c = 15000$, $c_p^{uso} = 10000$, $\mu_2 = .4$ and $\mu_1 = 1$. In this case, the long-run average cost in case of only corrective replacements is equal to 4285.71. In Table 1, we show the long-run average cost for various choices of λ , τ and c_p^{so} for three different policies: the optimal policy (a_{opt}), the policy that we replace at only the scheduled opportunities (a_{SO}), and finally, the policy under which we always replace at all opportunities (a_{alw}).

6. Conclusions and future work

In this paper, we derived the optimal replacement policy for a 3-state component degrading over time with corrective replacements at failures and preventive replacements at both scheduled and unscheduled opportunities. We also explicitly calculated the long-run average cost for certain time-dependent policies. In future work, we want to extend our analysis to models of components with more than three states, to models with additional features such as repairs, and to models in which the condition of the component is only partially observable.

Acknowledgments

The work of Sz. Kálosi is supported by the Data Science Flagship framework, a cooperation between the Eindhoven University of Technology and Philips. The work of S. Kapodistria is supported by an NWO Gravitation Project, NETWORKS, and a TKI Project, DAISY4OFFSHORE. The authors would like to thank M. Barbieri, J. Korst, and V. Pronk (all Philips Research) and O.J. Boxma (Eindhoven University of Technology) for their time and advice in the preparation of this work.

References

- Jardine, A.K.S., Lin, D., and Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7):1483–1510.
- Jardine, A.K.S. and Tsang, A.H.C. (2013). *Maintenance, Replacement, and Reliability: Theory and Applications*. CRC Press.
- Kobbacy, K.A.H. and Murthy, D.N.P. (2008). *Complex System Maintenance Handbook*. Springer.
- Manzini, R., Regattieri, A., Pham, H., and Ferrari, E. (2010). *Maintenance for Industrial Systems*. Springer.
- Prajapati, A., Bechtel, J., and Ganesan, S. (2012). Condition based maintenance: a survey. *Journal of Quality in Maintenance Engineering*, 18(4):384–400.
- Puterman, M.L. (2005). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley Series in Probability and Statistics.
- Ross, S.M. (1970). Average cost semi-Markov decision processes. *Journal of Applied Probability*, 7(3):649–656.
- Wang, H. (2002). A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, 139(3):469–489.
- Zhu, Q., Peng, H., and van Houtum, G.-J. (2016). An age-based maintenance policy using the opportunities of scheduled and unscheduled system downs. Beta report, Eindhoven University of Technology.

Competing deterioration processes modeling and RUL estimation based on Factorial Hidden Markov Models

T.T. LE, F. CHATELAIN and C. BERENGUER

Univ. Grenoble Alpes & CNRS, GIPSA-lab

{thanh-trung.le, florent.chatelain, christophe.berenguer}@gipsa-lab.grenoble-inp.fr

Abstract Deterioration modeling and Remaining Useful Life (RUL) estimation play an important role in implementing a predictive maintenance program since it helps to predict early the failure time caused by an incipient deterioration, and thus to provide sufficient time for maintenance crew to act before the actual failure eventually occurs. One well-identified challenge in deterioration modeling is that equipment does not always deteriorate following only one single mechanism, instead several deterioration modes can co-exist in competition due to the dynamics and variability of operating and environmental conditions. In this paper, we study the phenomenon where different modes co-evolve in parallel within a single component until one of them lead the component to fail. The observations contain hence information of the states of each mode, leading the inference problem being difficult. To deal with this problem, the Factorial Hidden Markov Model (FHMM) and the variational approximation technique are adapted. The associated RUL estimation method is also developed. To evaluate the performance of the proposed approach, a numerical study is given.

1. Introduction

Predictive Maintenance (PM) plays nowadays an important role in maintaining production equipment thanks to its ability to eliminate unnecessary scheduled preventive maintenance operations as well as to avoid unforeseen failures; hence it can significantly reduce the maintenance cost Jardine et al. (2006). Within a PM framework, the RUL estimation is a key enabling task since it could provide sufficient time for maintenance engineers to schedule an intervention or to acquire replacement components before an actual failure. As pointed out in Gorjian et al. (2010), evaluation of the current equipment's health state is an essential task in estimating its RUL in practical applications. To this end, deterioration phenomena of the equipment should be carefully investigated and properly modeled.

In the literature, stochastic models are often studied to model deterioration processes thanks to its capacity of taking into account the stochastic nature of the deterioration phenomena. Depending on the types of its states space, one can divide the stochastic models into two main classes: continuous or discrete. Lévy process Van Noortwijk (2009), general path model Lu & Meeker (1993) are examples of the first class. Regarding the continuous-state processes, we can list here Markovian-based models, such as Hidden Markov Models (HMMs) Rabiner (1989), Hidden semi-Markov models (HsMM) Dong & He (2007). Compared to the first one, the second class models the health states of equipment by a finite number of discrete states which could be easier for maintenance engineers to interpret the obtained results. A detailed review of deterioration models used in reliability analysis can be found in Gorjian et al. (2010).

While modeling the deterioration processes, a well identified difficulty is to deal with the co-existence of multiple modes of deterioration Compare et al. (2015); Le et al. (2016). These different mechanisms, coexisting even within a single component, can lead to different temporal evolution behaviors of deterioration indicators. An example of such phenomenon can be found in a rolling element bearing, an essential component of every rotating machinery in industry. In effect, under operation, point defects such as cracks, pits and spalls appearing on different constituent elements (i.e. outer race, inner race, cage, etc.) can lead to different deterioration dynamics of the bearing Heng (2009). Taking this phenomenon into account is hence crucial both for a comprehensive assessment of the health state of equipment as well as for an accurate RUL estimation. To tackle this problem, the concept of multi-branch models has been proposed recently by the current authors in Le et al. (2016, 2015b). As Markov-based models, the multi-branch methods showed promising results in RUL estimation compared to the ones obtained with the standard "mono-branch" models especially when multiple deterioration modes co-exist. Nevertheless, the works in Le et al. (2016) considered only the modes competition in an incubation period, meaning that once one of them has already initiated, the equipment cannot change

to the others. In Le et al. (2015a), the modes transitions were allowed but the equipment can be only in one of them at a time.

In this paper, we are interested in modeling the deterioration phenomenon where several modes coexist in parallel. Specifically, different deterioration modes evolve independently and all contribute to observed default symptoms. It is sufficient that only one of these modes reaches its own failure state can lead to the breakdown of the equipment. The mixed contribution of the modes at the observation level leads to difficulties in diagnosing the equipment's health state as well as in predicting its RUL. To overcome these problems, we propose in this paper a diagnostics/prognostics framework based on the factorial hidden Markov model (FHMM) Ghahramani & Jordan (1997). The remainder of the paper is organized as follows. Section 2 is devoted to review some mathematical background of the FHMM model and explain how to estimate its model parameters via an approximated method. The application of the FHMM model for the diagnostics and prognostics of a bearing is investigated in Section 3. Finally, conclusion follows in Section 4.

2. Factorial Hidden Markov Model

2.1. Model structure

The Factorial Hidden Markov Model (FHMM) is an extension of the HMM one with distributed state representation which is a particular class of probabilistic graphical model Ghahramani & Jordan (1997). The model consists of several branches in parallel in which each one can be considered as a single Markov chain. At any instants, the distributed states all contribute to the observations as shown in Figure 1.

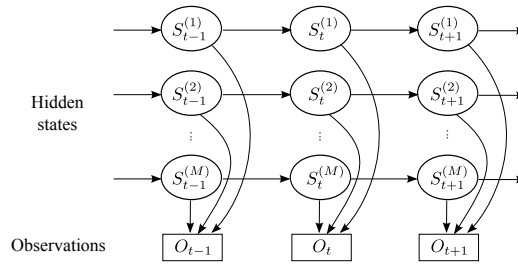


Figure 1. Factorial Hidden Markov Model

Similar to the standard HMM models, states transitions within a branch m of the FHMM model is characterized by an initial state distribution $\pi^{(m)}$ and a state transition matrix $P^{(m)}$ Rabiner (1989) where the superscript is used to denote the corresponding branch in this paper. A combination of the states in all constituent branches at a given time t is called the meta-state, that is $S_t = (S_t^{(1)}, S_t^{(2)}, \dots, S_t^{(M)})$ where M is the number of branches. Furthermore, the states within the branch m can take one of $K^{(m)}$ discrete values, i.e. $S_t^{(m)} \in \{1, 2, \dots, K^{(m)}\}$. For the seek of simplicity, we assume in this paper that $K^{(m)} = K, \forall m$. The generation to the case of different $K^{(m)}$ values is straightforward Ghahramani & Jordan (1997).

In the FHMM model, the observation emitted at a given time depends on the individual states within the branches at that time. Its density can be modeled as a function of the ones of the branch states:

$$p(O_t | S_t) = f \left[P(O_t | S_t^{(1)}), P(O_t | S_t^{(2)}), \dots, P(O_t | S_t^{(M)}) \right] \quad (1)$$

The precise nature of the function f depends on the proportions to which the observations from the individual components are mixed in the current one. The contribution of each component could depend on several factors and is difficult to be determined precisely Reyes-Gomez et al. (2003). For continuous observations, a simple form for this dependence is linear Gaussian Ghahramani & Jordan (1997). That is, the observation O_t is a Gaussian random vector whose mean is a linear function of the component ones:

$$\mu_t = \sum_{m=1}^M w_\mu^{(m)} \mu_t^{(m)} \quad (2)$$

where $\mu_t^{(m)}$ represents the D dimensional mean vector of the Gaussian distribution corresponding to the component state $S_t^{(m)}$ and $w_\mu^{(m)}$ is a $D \times D$ weighting matrix.

Concerning the covariance parameter, there exists two main possibilities in the literature : global or composed covariance approaches Reyes-Gomez et al. (2003). It should be noted that the composed covariance could lead the model learning procedure being more complex than in the global covariance case since it introduces more free parameters to be estimated. For the seek of simplicity, only the global covariance matrix is considered in this paper, meaning that all factorial states have in common a covariance matrix C .

2.2. Parameters estimation

Denote $\Theta = \left\{ \pi^{(m)}, P^{(m)}, w_{\mu}^{(m)}, C \right\}_{m=1:M}$ the parameters vector of the FHMM model. Since the states are hidden, the standard expectation-maximization (EM) algorithm Dempster et al. (1977) appears to be a suitable tool for parameters estimation. However, as pointed out in Ghahramani & Jordan (1997), the expectation (E) step is computationally intractable in this case. Therefore we adapt in this section the variational approximation technique proposed by Ghahramani & Jordan (1997) to overcome this problem.

The basic idea of the variational technique is to approximate the posterior distribution over the hidden variables $P(S_{1:T} | O_{1:T})$ by a tractable distribution $Q(S_{1:T})$ where $S_{1:T}$ and $O_{1:T}$ denote the sequence of meta-states and observations from time 1 to T . More specifically, at any k th iteration of the standard EM algorithm, the Jensen's inequality can be applied in computing the expected complete data log-likelihood as follows:

$$\begin{aligned} \mathcal{Q}(\Theta | \Theta^{k-1}) &= \log \sum_{S_{1:T}} P(S_{1:T}, O_{1:T} | \Theta) = \log \sum_{S_{1:T}} Q(S_{1:T}) \left[\frac{P(S_{1:T}, O_{1:T} | \Theta)}{Q(S_{1:T})} \right] \\ &\geq \sum_{S_{1:T}} Q(S_{1:T}) \log \left[\frac{P(S_{1:T}, O_{1:T} | \Theta)}{Q(S_{1:T})} \right] = \mathcal{F}(Q, \Theta) \end{aligned} \quad (3)$$

This inequality provides a lower bound on the log-likelihood that can be used to obtain an efficient learning algorithm. The difference between $\mathcal{Q}$ and $\mathcal{F}$ is given by the Kullback-Leibler divergence:

$$KL(Q||P) = \sum_{S_{1:T}} Q(S_{1:T}) \log \left[\frac{Q(S_{1:T})}{P(S_{1:T} | O_{1:T}, \Theta^{new})} \right] \quad (4)$$

By minimizing this divergence, the log-likelihood is assured to increase until it reaches a (local) maximum.

It should be noted that the function Q must be chosen to be tractable. Depending on the form of this function, two variational techniques were proposed: completely factorized and structured variational inference Ghahramani & Jordan (1997). Compared to the first one which factors the posterior probability such that all the state variables are statistically independent, the second technique can preserve more the probabilistic structure of the original FHMM model and hence will be used in this paper.

Based on this approximation technique, the parameters vector Θ is computed in the maximization (M) step. More details about the calculation formulas can be found in Ghahramani & Jordan (1997).

2.3. State estimation and RUL prediction

Suppose that we have trained the FHMM model in an offline phase from the historical until-failure data. We move now to an online phase where the deterioration signals of a monitored system are acquired in real time via a condition monitoring system. The objective is to estimate the actual states of all the deterioration modes at a given time and then predict the system's RUL.

The estimation of the deterioration states given the observations in case of the FHMM model is itself a complex task since it can be referred as the source separation problem found in the speech recognition community Reyes-Gomez et al. (2003). Several methods have been developed in the literature for accomplishing this problem. One of the most simple solutions is to transform the FHMM model into a meta-HMM one with K^M "meta-states" and then applied the Viterbi algorithm Rabiner (1989) in a standard way. Although this method leads to more computation time compared to another ones, its performance is still acceptable in several practical applications since only the maximization is retained after each recursive iteration.

This method allows us to estimate the most likely sequence of meta-state given a sequence of observations. The corresponding state sequences of each individual branch of the FHMM model is then obtained by decomposing it into a base-M system. After that, the last state of each decomposed sequence can be considered as the

current health state of the deterioration mode corresponding to that sequence. Given the estimated states, the RUL can be defined as follows:

$$RUL = \min_k RUL^{(k)} \quad (5)$$

where $RUL^{(k)}$ is the remaining time for the mode k to reach the final state from its current state. In other words, $RUL^{(k)}$ is referred as the RUL of a standard HMM model corresponding to the k th branch of the FHMM model. The RUL probability density function (pdf) calculation for a standard HMM model for the discrete time case can be found in Le et al. (2016).

Given the individual RUL density functions, the cumulative distribution function (cdf) of the overall RUL of the FHMM model can be obtained by:

$$CDF_{RUL}(x) = 1 - \mathbb{P}(RUL^{(1)} > x, RUL^{(2)} > x, \dots, RUL^{(M)} > x) = 1 - \prod_{k=1}^M (CDF_{RUL}^{(k)}(x)) \quad (6)$$

3. Numerical study

In this section, the diagnostics and prognostics performances of the FHMM model are evaluated through the case of a defected rolling element bearing with multiple point defects. The bearing is chosen because it is an essential component for all rotating machinery found in industry. Firstly, we simulate vibration signals generated by a bearing in case that several point defects coexist within it. The kurtosis features are then extracted and used as observations for training the FHMM model. The RUL estimation performance of the learned FHMM model is compared with the one obtained by a standard HMM model.

3.1. Multiple point defects simulation with severity evolution

McFadden & Smith (1985) proposed a simple mathematical model for simulating the vibration signals produced by a bearing with multiple point defects. By assuming that the bearing operates under constant radial load, the vibration response $v(t)$ can be modeled as the product of repetitious impulses $d(t)$ at the fault frequency with the distribution of load $q(t)$, convolved with an exponential decay $h(t)$ due to damping:

$$v(t) = [d(t) \cdot q(t)] * h(t) + n(t) \quad (7)$$

where the notation $*$ represents the convolution operation and $n(t)$ are the noises added to corrupt the signal.

Figure 2a shows the simulated raw vibration signals of the bearing in case that two point defects co-exist, one is on the inner race and the other on the outer race. The co-existence of these two defects can be verified on the frequency spectrum obtained through an envelope analysis (c.f. Figure 2b).

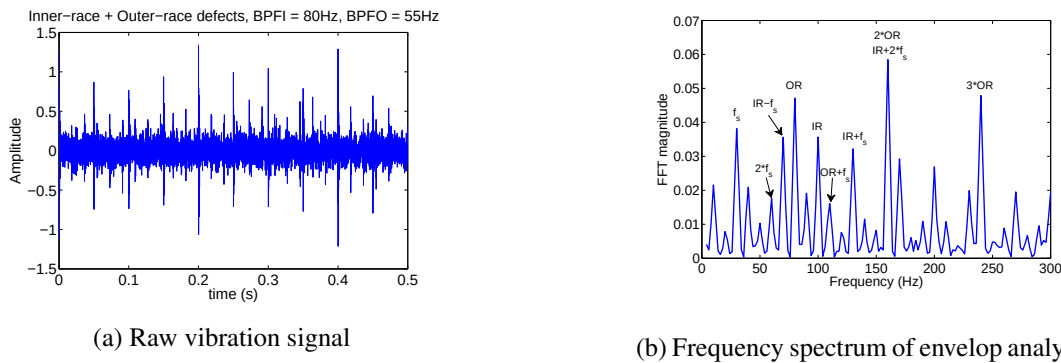


Figure 2. Multiple point defects bearing simulation

This model although represents well the vibration signals of a defected bearing, it does not however take into account the temporal evolution of the defaults. In the literature, vibration data are usually acquired periodically and temporal features such as the kurtosis are extracted and plotted in order to study the evolution trend of fault severity over time for the bearing prognostics purpose. To simulate deterioration processes, the above defective bearing vibration signals are repeatedly generated with progressive impulse amplitudes $d(t)$. Each repetition corresponds to one data acquisition time in practice. In this study, we suppose that the impulse amplitudes $d(t)$

generated by each point defect increase over time following a homogeneous Gamma process. A random delay following a Poisson distribution is also added in order to represent the bias between the appearance times of the two defects. The kurtosis feature is used to represent the overall defect severity of the bearing. Finally, the data generation is stopped once the extracted kurtosis value reaches for the first time a critical threshold $L = 10$. Beyond this boundary, the bearing is considered to be failed.

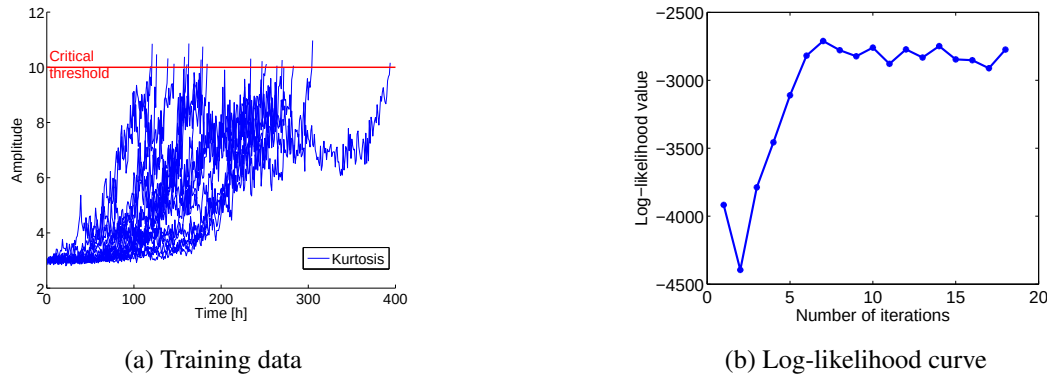


Figure 3. Training data and learned log-likelihood curve

By repeating this procedure several times, different kurtosis evolution curves can be obtained as represented in Figure 3a. These values are considered as the training data for estimating the parameters of the FHMM model. Figure 3b shows the convergence of the log-likelihood curve which demonstrates the effectiveness of the variational approximation technique in training the FHMM model.

3.2. FHMM vs HMM

In studying the FHMM model, an interesting question is to compare it with a “standard” HMM model in term of RUL estimation performance. To this end, we firstly train the two models by the same training data set. The branches of the FHMM model are supposed to have the same number of discrete states as in the HMM one. This number can be determined by using the Bayesian Information Criterion (BIC) as presented in Le et al. (2016).

The next step is to use the two learned models to estimate the RUL of same test data set. The root mean squared errors (RMSE) is used for the comparison purpose. Since the prognostics performance depends on the instant at which the RUL estimation is carried out, we further define a critical level at which the prediction is triggered for all the test cases. In other words, the diagnostic and prognostic tasks for every test cases are started once its corresponding kurtosis feature reaches this level for the first time. Figure 4 shows the RMSE values obtained by the two models as the function of the triggered level.

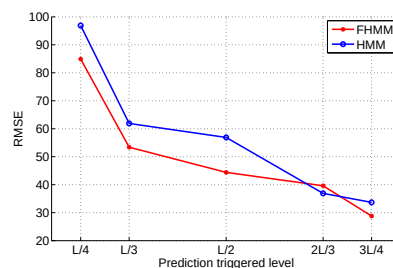


Figure 4. RMSE values at different prediction triggered levels

We can see that the RUL estimation errors for the both models decrease with the increment of the triggered level. This is due to the fact that the latter the prediction is performed, the more information become available and hence leading to the better RUL estimation results. Furthermore, compared to the standard HMM model, the FHMM one shows the better performance in terms of RMSE criterion, especially when one want to trigger the prediction earlier. This can be explained by the fact that the FHMM model helps to take into account

the coexistence of different deterioration modes at early stage of deterioration process. More lately, one of the deterioration modes may already dominate the others, leading to the similar RMSE results between the FHMM and the HMM models.

4. Conclusion

The present paper proposes the use of the factorial Hidden Markov Model (FHMM) in order to deal with the deterioration modeling problem in case that several competing modes co-exist within a single component. The variational approximation technique is adapted to overcome the model parameters learning problems. In terms of diagnostics, this model helps to evaluate the actual deterioration states within every existed modes. The RUL of the component is then estimated as the minimum time for one of these modes to reach its own failure state. The numerical results show that by taking into account the coexistence of several deterioration modes, the FHMM gives better performance in RUL estimation in comparison with the traditional HMM model.

Future research will be focused on the extension of the FHMM model to represent the dependency between the different deterioration modes. The proposed method should also be validated on real world systems data.

References

- M. Compare, et al. (2015). 'Interacting multiple-models, state augmented Particle Filtering for fault diagnostics'. *Probabilistic Engineering Mechanics* **40**:12–24.
- A. P. Dempster, et al. (1977). 'Maximum likelihood from incomplete data via the EM algorithm'. *Journal of the royal statistical society. Series B (methodological)* pp. 1–38.
- M. Dong & D. He (2007). 'A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology'. *Mechanical Systems and Signal Processing* **21**(5):2248–2266.
- Z. Ghahramani & M. I. Jordan (1997). 'Factorial hidden Markov models'. *Machine learning* **29**(2-3):245–273.
- N. Gorjian, et al. (2010). 'A review on degradation models in reliability analysis'. In *Engineering Asset Lifecycle Management*, pp. 369–384. Springer.
- A. S. Y. Heng (2009). *Intelligent prognostics of machinery health utilising suspended condition monitoring data*. Ph.D. thesis, Queensland University of Technology.
- A. K. Jardine, et al. (2006). 'A review on machinery diagnostics and prognostics implementing condition-based maintenance'. *Mechanical systems and signal processing* **20**(7):1483–1510.
- T. T. Le, et al. (2015a). 'Prognosis based on Multi-branch Hidden semi-Markov Models: A case study'. In *Proc. Fault Detection, Supervision and Safety of Technical Processes Conference, Paris, France, September 2015*, pp. 91–96.
- T. T. Le, et al. (2015b). 'Jump Markov Linear Systems for deterioration modeling and Remaining Useful Life estimation'. In *Proc. European Safety and Reliability Conference, Zurich, Switzerland, September 2015*, pp. 1287–1295.
- T. T. Le, et al. (2016). 'Multi-branch hidden Markov models for remaining useful life estimation of systems under multiple deterioration modes'. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*.
- C. J. Lu & W. O. Meeker (1993). 'Using degradation measures to estimate a time-to-failure distribution'. *Technometrics* **35**(2):161–174.
- P. McFadden & J. Smith (1985). 'The vibration produced by multiple point defects in a rolling element bearing'. *Journal of sound and vibration* **98**(2):263–273.
- L. R. Rabiner (1989). 'A tutorial on hidden Markov models and selected applications in speech recognition'. *Proceedings of the IEEE* **77**(2):257–286.
- M. J. Reyes-Gomez, et al. (2003). 'Multi-channel source separation by factorial HMMs'. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, vol. 1, pp. I–664. IEEE.
- J. Van Noortwijk (2009). 'A survey of the application of gamma processes in maintenance'. *Reliability Engineering & System Safety* **94**(1):2–21.

Life Model of Aluminum Alloy Crevice Corrosion based on Mechanical Property under Coupling Environment Conditions

LI Zheng*, LEI Peng*, Bai Fengyu*, Wang Jian**

* Northwestern Polytechnical University, Xi'an, 710072, China, email address: lizhchi@nwpu.edu.cn

** Kingston University, London SW15 3DW, UK, email address: j.wang@kingston.ac.uk

Abstract. The crevice (characterized by crevice width C) of LY-12 sheets jointed by fastening bolt could be exposed to corrosion during their whole life cycle in an environment with temperature and saline solution. There appears to be lack of life model of aluminum alloy with consideration of this type of corrosion that can be used to predict the deterioration of mechanical properties. In the present paper, a mathematical model has been developed to evaluate the longevity of the aluminum alloy. Crevice corrosion is modeled by characterizing the relationship between mechanical properties of LY-12 connecting structure and corrosion depths based on corrosion experiments with different crevice width, temperature and saline solution. The dependent variable for the conservation rate of mechanical property called the reliability parameter (R), was used to create a comprehensive relationally statistic model, which are functions of unsteady (times' function) independent variables temperature (T), saline solution concentration (c) and crevice width (C). The service life can be predicted by the life model at a given level of deterioration of mechanical property due to different environment conditions and different structure crevice width. It is essential for aircraft safety to predict the degenerated life and reduced reliability under real corrosion environments.

1. Introduction

Serious crevice corrosion of aircraft structure often occurs during the actual service at different component connecting locations. Metal connection structure potentially exposes to crack initiation and extension, which leads to aircraft structure failure. Those kinds of structural failures affect the structural integrity and reliability of the plane and might directly cause sudden and often catastrophic failure because that they would not be able to detected effectively and promptly^[1]. It is widely accepted that the life of aluminum alloy crevice corrosion is important for predicting the propagation of corrosion damage in order to assess the risk of failure.

Many research efforts have been directed to predict to model the alloy crevice corrosion life. F.M. Song^[2] developed a model to predict the chemistry, corrosion potential and the proportion of pipeline steel in a crevice formed when a coating disbands from a pipe surface. Wen Sun et al^[3] proposed an arbitrary Lagrangian-Eulerian (ALE) model that can simulate the transient potential distribution, the dynamic concentration profiles and the time-dependent geometry deformation for a corroding crevice. O.K. Morachkovskii and Yu.V. Romashov^[4] proposed a mathematical model of crack propagation caused by the corrosion cracking of steels with an aim of predicting the service life of structural elements. J.C. Walton et al^[5] presented a transient mathematical/numerical model to simulate development processes of crevice. Shu-Xin Li et al^[6] presented a method for predicting the corrosion remaining life of underground pipelines. Hyun-Young Chan et al^[7] developed a mathematical model to evaluate the transport processes and chemical reactions during the initiation stage of crevice corrosion. These models would be helpful in elucidating the mechanism of crevice corrosion and other forms of related localised corrosion. But there lacks a life model of crevice corrosion that can predict the deterioration of mechanical property as the prediction of reliability and remaining life are of great importance in practical scenarios.

As is well-known, the corrosion pits will lead to mechanical property deterioration. The deterioration of mechanical property can be described by conservation rate (R) and depth of the corrosion pits can be used to describe the degree of environmental corrosion. Based on the example of Aluminum alloy (LY-12) structures with common fasteners of bolt connection commonly used on the plane, the paper aims to develop a mathematical model to evaluate the longevity of aluminum alloy crevice corrosion and to characterize the relationship between mechanical property and corrosion pit depth based on experiments.

2. Crevice corrosion and mechanical property test

2.1 Preparation of the Specimens

Specimens to be tested in the present case are aluminum alloy structures with bolt connections. The aluminum sheets are marked as LY-12CZ. The aluminum sheets are processed to a dog-bone-shape, with a thickness of 3mm and 180mm length as shown in Figure 1. Tapping is made at the center of each aluminum sheet, with a boresize of 4mm. The aviation general 30CrMnSi bolts ($\Phi 4\text{mm}$) are chosen as the attachment bolts. The corrosion region covering the area 32mm (L) by 26mm (W) around the center hole, and the remaining part of the specimen was sealed by wax. The crevice widths for the present test case are 0.02mm, 0.06mm and 0.1mm.

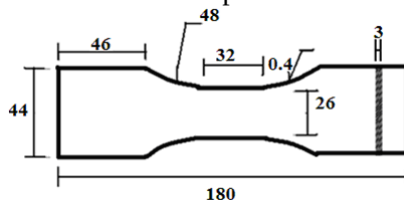


Figure1.Aluminum sheet used in the test

2.2 Experiment

The test is divided into two parts. A typical environment immersion corrosion test is the first part. The testing environment and crevice width are shown in Table 1. Test of the mechanical properties of aluminum sheet after corrosion has taken place forms the second part of the experiment. Both initial samples and all samples after corrosion tests are subjected to astatic tensile test with the loading rate of 1000N/S.

Table1. Cases of crevice corrosion test of LY-12CZ connection structures

crevice width (mm)	Sample Numbers	Temperature	PH	NaCl Solution (%)	Soaking time (day)	Code
0.02, 0.06, 0.1	3×5	25℃	4.2	3.50	30	Env1
0.02, 0.06, 0.1	3×5	40℃	4.2	3.50	20	Env2
0.02, 0.06, 0.1	3×3	40℃	4.2	3.50	15	Env3
0.02, 0.06, 0.1	3×5	25℃	4.2	3.00	30	Env4
0.02, 0.06, 0.1	3×4	40℃	4.2	3.00	30	Env5

3 Relationship analysis between corrosion pits and mechanical property

3.1 Theoretical model

As ultimate stress mainly affected by cross-sectional area during static tension^[8], it can get as follows to the specimens

$$\delta = \frac{F}{b_1 l + b_2(l - D)}. \quad (1)$$

Where,

δ --Stress; F --ultimate load; l --thickness of aluminium sheet; D --Depth of corrosion pit;

b_1, b_2 --Width sum of no corrosion area and of corrosion area.

As definition of conservation rate of mechanical property is:

$$R = \frac{E_t}{E_0}. \quad (2)$$

Where,

R --conservation rate of mechanical property;

E_t --average value of mechanical property after crevice corrosion;

E_0 --average value of mechanical property before crevice corrosion.

The stresses are taken as the mechanical property of testing samples. As $(l-D) \rightarrow l$, the simplified formula of conservation rate of mechanical property can be expressed as

$$R = \hbar D + \varepsilon. \quad (3)$$

Where,

R --conservation rate of mechanical property; D --Depth of corrosion pits, unit μm .

3.2 Data fitting and validation

A linear relationship (shown as Figure 2) exists between R and D and from the curve fitting of testing data (Maximum deviation is 1.38%), functional relation expressed as :

$$R = -0.014D + 1.188. \quad (4)$$

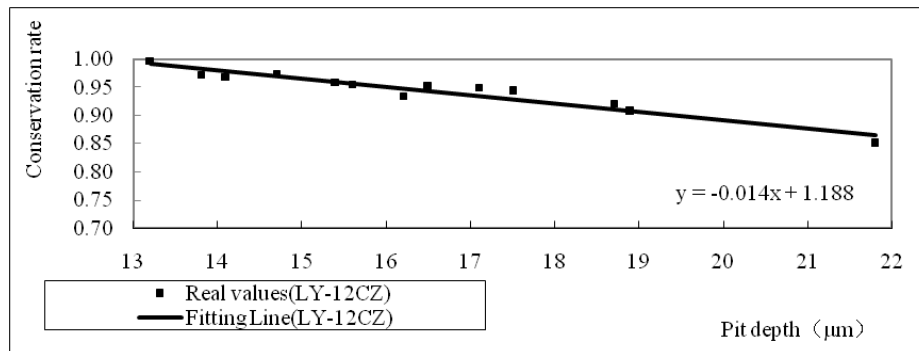


Figure 2. Fitting line between depth of corrosion pits and conservation rate

Therefore, the depth of corrosion pits should be a parameter that could link the environmental conditions and the mechanical property. Based on the depth of corrosion pits, the deterioration of the mechanical property can be analyzed under any given environment.

4 Influence analysis of mechanical property by structure and environment

Because the variables (temperature, saline solution concentration and) are not independent, it can be studied by orthogonal experimental design at different corrosion periods on the hypothesis that only one variable can be changed. So, the influence of any variable is analyzed by relevant experimental data, such as the influence of crevice width by any testing environment, the influence of temperature/crevice width by Env1 and Env2, and the influence of saline solution/crevice width by Env1 and Env4. Then, the function relationship can be concluded by variable substitution.

4.1 Effect on crevice width for conservation rate

It can be derived that the variation tendency of pits depth decreases gradually and mechanical property increases gradually accordingly from 0.02mm, 0.06mm to 0.1mm based on experimental data. It can be thought as reasonable to assume that the variation tendency is similar for all the other of pit depth that a decrease in depth causes an increase in mechanical property from depth of 0.02mm to 0.1mm. The quadratic equation can be obtained by polynomial interpolation between gap size and pits depth on every testing environment without testing the samples. The general formula can be expressed as

$$D = mC^2 + nC + q. \quad (5)$$

Where,

D --Depth of corrosion pit, μm ; C --Crevice width, mm ; m, n, q --Coefficient.

So,

$$R = h(mC^2 + nC + q) + \varepsilon. \quad (6)$$

4.2 Effect on temperature/crevice width for conservation rate at any time

(1) Data analysis in "Env2" (40 °C, 3.5% sodium chloride)

A power function can be used to describe the changing of pits depth with the time^[9]

$$D|_t = A\tilde{t}^\alpha. \quad (7)$$

Where,

$D|_t$ --Pitting depth at time t ; A --Parameter; α --Coefficient.

When $x_i = \ln t_i$, $y_i = \ln D_i$, then parameters can be derived in the same concentration of sodium chloride from analysis of the testing data on (x_i, y_i) ($i=1, 2, \dots, n$), shown as Table 2.

Table 2. Regression parameter values of the power function

Crevice width	Temperature	Sample Number	$\bar{x}$	$\bar{y}$	$\ln \hat{A}$	$\hat{\sigma}$	$\hat{\alpha}$
---------------	-------------	---------------	-----------	-----------	---------------	----------------	----------------

Crevice width	Temperature	Sample Number	$\bar{x}$	$\bar{y}$	$\ln \hat{A}$	$\hat{\sigma}$	$\hat{\alpha}$
0.02 mm	40	8	2.89	2.91	2.005	0.227	0.313
	25	4	3.39	2.91	1.849		
0.06 mm	40	7	2.87	2.73	1.843	0.283	0.309
	25	5	3.39	2.75	1.702		
0.1 mm	40	8	2.89	2.73	1.869	0.289	0.298
	25	5	3.39	2.76	1.750		

Combining simultaneous equations (4),(5)and(7)when averaging $\hat{\alpha} = 0.3$, the general model of R and C in “Env2” at any time can be given by

$$R = 45.048t^{0.3} \cdot C^2 - 0.6.679t^{0.3} \cdot C + 0.091t^{0.3} + 1.265$$

(2) Data analysis in “Env1 ”(25 °C,3.5% sodium chloride)

According to the definition of accelerated corrosion factor, when the corrosion volume is equal in the natural environment and laboratory accelerated environment, it can be get

$$K_p = \frac{\tilde{t}}{t} \Big| D_p = \tilde{D}_p. \quad (8)$$

Then, at any same time t in the two environments, the formula is

$$D_p \Big| t = K^a \cdot \tilde{D}_p \Big| t. \quad (9)$$

In a similar way, when $\hat{\alpha} = 0.3$, the general model of R and C in Environment 1 at any time is

$$R = -6.916t^{0.3} \cdot C^2 + 0.975t^{0.3} \cdot C - 0.047t^{0.3} + 1.265$$

(3) Universal law at any temperature

Integrating the formula of volume growth rate of corrosion pits and setting the shape of corrosion pits hemispheroid, that is

$$\frac{dV}{dt} = \frac{MI_{p_0}(k)}{nF\rho} \left(-\frac{E_a}{R_0T} \right), \quad V = \frac{2}{3}D^3\pi$$

Then, a mathematical formula can be used to obtain the pits depth at any time and temperature.

$$D^3 = D_0^3 + \frac{3}{2\pi} \frac{MI_{p_0}(k)}{nF\rho} \cdot \exp\left(-\frac{E_a}{3R_0T}\right) \cdot t. \quad (10)$$

Where,

D_0 --The initial corrosion pit depth, μm ; V --The volume of corrosion pit;

I_{p_0} --Electric current coefficient of corrosion pit; E_a --Activation energy;

T --Temperature; K ; ρ --Density; n --Chemical valence;

M --Atomic weight; F --faraday constant, $F=96514 C \cdot mol^{-1}$; t --Time, ; day;

R_0 --perfect gas constant, $R_0=8.314 J \cdot (K \cdot mol)^{-1}$.

Therefore, universal law of the function relational expression of C - t - T - R at any temperature satisfies the following form by combining the formula (10) and (3).

$$R = \beta \cdot t^{\frac{1}{3}} \cdot (C - \theta)^2 \cdot \exp\left(-\frac{E_a}{3R_0T}\right) + \varepsilon. \quad (11)$$

Where, $\hat{\beta}$, θ and ε are constant coefficient.

4.3 Effect on saline solution /crevice width for conservation rate at any time

Corrosion pits tend to grow in the aluminum plate surface when metallic aluminium dissolve and generate insoluble hydroxide and free acid during the process of crevice corrosion.

The volume of consumption of metallic aluminium, that is the corrosion pit size, can be regarded as the criteria of measurement for product of aluminum hydroxide because the quantity of per unit consumption of the metal will translate into an equal amount of hydroxide.

To compare two different kinds of corrosive environment, the specimens were soaked for thirty days in 3% and 3.5% sodium chloride solution respectively at 25(Env1 and Env4), the main difference being the chloride ion concentration of the reactants. Then, from the chemical reaction rate

$$\tilde{R} = v(T) \cdot [E]^m [F]^n$$

It can be concluded that

$$\frac{D_1^3}{D_4^3} = \left[\frac{[Cl^-]_1}{[Cl^-]_4} \right]^m$$

Where,

$[E], [F]$ -- The concentration of the reactants; $\tilde{R}$ -- Chemical reaction rate;

$v(T)$ -- constant of chemical reaction rate (related with temperature);

D_4, D_1 -- Corrosion pit depth in 3%, 3.5% sodium chloride solution;

$[Cl^-]_4, [Cl^-]_1$ -- Chloride ion concentration of the reactants in 3%, 3.5% sodium chloride solution.

The relationship between corrosion pit depth and sodium chloride solution concentration can be described as follows:

$$\ln D = \ln D_4 + \frac{m}{3} \ln \frac{c}{3.5}. \quad (12)$$

4.4 Theoretical model of C-T-c-R on coupling environments

(1) The general functional form

By the analysis of the expression presented in the preceding section, it can be concluded that the general functional form of the relationship between the mechanical property, crevice width, temperature and Sodium chloride solution concentration at any time (C-t-T-c-D) is shown as follows.

$$D = \beta \cdot \left(\frac{3}{2\pi} \frac{M I_{P_0}}{n F \rho} \right)^{\frac{1}{3}} \cdot t^\alpha \cdot (C - \theta)^2 \cdot \exp\left(-\frac{E_a}{3R_0 T}\right) \cdot \left(\frac{c}{3.5}\right)^{\frac{m}{3}}. \quad (13)$$

$$R = \hat{\beta} \cdot \left(\frac{3}{2\pi} \frac{M I_{P_0}}{n F \rho} \right)^{\frac{1}{3}} \cdot t^\alpha \cdot (C - \theta)^2 \cdot \exp\left(-\frac{E_a}{3R_0 T}\right) \cdot \left(\frac{c}{3.5}\right)^{\frac{m}{3}} + \varepsilon \sim \frac{t^\alpha c^{\frac{m}{3}} (C - \theta)^2}{e^{\frac{1}{T}}}. \quad (14)$$

(2) Verification and comparison with experimental data

According to the former formula by the relevant experimental data from “Env1” to “Env4”, these parameters can be obtained as follows: $\alpha = 1/3$; $\theta = 0.965$; $m = 0.802$; $\varepsilon = 1.265$.

Then,

$$D = 2947 \cdot t^{\frac{1}{3}} \cdot (0.965 - C)^2 \cdot \exp\left(-\frac{1844}{T}\right) \cdot \left(\frac{c}{3.5}\right)^{\frac{0.802}{3}}.$$

$$R = -57.665 \cdot t^{\frac{1}{3}} \cdot (0.965 - C)^2 \cdot \exp\left(-\frac{1844}{T}\right) \cdot \left(\frac{c}{3.5}\right)^{\frac{0.802}{3}} + 1.265.$$

Using the formula, conservation rate can be worked out at any experimental period in “Env5”. Comparison of the results for the calculated value and the experimentally obtained are given in Table 3. Based on the results it can be proved that the model is reasonable due to the low margin of error (no more than 3.37%).

Table 3. Verification data R in “Env5”

Crevice width (mm)	conservation rate R		
	calculated value	testing value	Error
0.02	0.819	0.830	1.36%
0.06	0.856	0.841	1.74%
0.1	0.891	0.862	3.37%

5 The corrosion time estimates in different corrosion environments

By using the universal relation model of corrosion rate and environment factors, the corrosion time can be estimated maintaining a certain value of conservation rate R in any concrete corrosion environment. As an example in two experimental environments, the results of R as 90%, 80%, 70% and 60% respectively are listed in Table 4 below.

Table 4. Corrosion time estimation at different R in specific environment

Soap environment	Crevice width(mm)	Time (day)			
		R=90%	R=80%	R=70%	R=60%
40°C, 3.5%	0.02	25	60	135	270
	0.06	30	75	180	345
	0.1	35	105	240	435
25°C, 3%	0.02	60	165	375	690
	0.06	90	225	495	915
	0.1	105	285	660	1140

From Table 4, it is clearly demonstrated that a smaller the degree of corrosion is found with an increase in the gap size in the very same test environment. The corrosion is most serious at around 0.02 mm gap size. The corrosion degree of the structure is significantly increased due to a temperature rise rather than with concentration of salt solution rise.

6. Conclusions

- (1) The relationship between the conservation rate and crevice width/temperature/concentration of salt solution tends to have a specific relation at any time of aluminum alloy connection structure in any corrosive environment. The dependent variable (R) is a function of unsteady (times' function) independent variables such as temperature (T), salt solution concentration (c) and crevice width (C). It can be used as a life time model to measure the degradation of key performance indicators with coupling factors.

$$R = \hat{\beta} \cdot \left(\frac{3}{2\pi} \frac{MI_{R_0}}{nF\rho} \right)^{\frac{1}{3}} \cdot t^{\alpha} \cdot (C - \theta)^2 \cdot \exp\left(-\frac{E_a}{3R_0T}\right) \cdot \left(\frac{c}{3.5}\right)^{\frac{m}{3}} + \varepsilon \sim \frac{t^{\alpha} c^{\frac{m}{3}} (C - \theta)^2}{e^{\frac{1}{T}}}$$

- (2) Within the scope of the permissible engineering tolerance level, we can predict time needed for a certain degree of mechanical performance degradation of air bearing aluminium alloy components through life model as long as there is a knowledge of the soaking corrosion environment conditions. The mathematical model presented in the current paper can reasonably predict the degradation as well the corrosion life of the system. As a result, it will lower the processes that are involved in practical corrosion life testing. The mathematical model presents a more practical method of calculating the corrosion life as opposed to the practice of visually inspecting the degree of corrosion.

Acknowledgement

The authors are grateful for the support provided by the Aeronautical Science Foundation of China (No. 20120253004).

References

- [1] George Engelhardt, Digby D. Macdonald. Estimation of corrosion cavity growth rate for predicting system service life[J]. Corrosion Science 46 (2004) pp 1159-1187.
- [2] F.M. Song. A mathematical model developed to predict the chemistry and corrosion rate in a crevice of variable gap[J]. Electrochimica Acta, 56 (2011) pp 6789 -6803.
- [3] Wen Sun, Lida Wang, Tingting Wu, Guichang Liu. An arbitrary Lagrangian-Eulerian model for modelling the time-dependent evolution of crevice corrosion[J]. Corrosion Science, 78(2014) pp 233-243.
- [4] O. K. Morachkovskii, Yu. V. Romashov. Continual model of propagation of corrosion cracks for the evaluation the service life of structures[J]. Materials Science, Vol. 46, No. 2 (2010) pp 254-259.
- [5] Jun Hu, Tian Yangyang, Haipeng Teng, Lijun Yu, Maosheng Zheng. The probabilistic life time prediction model of oil pipeline due to local corrosion crack[J]. Theoretical and Applied Fracture Mechanics 70 (2014) pp 10-18.
- [6] Shu-Xin Li, Shu-Rong Yu, Hai-Long Zeng, Jian-Hua Li, Rui Liang. Predicting corrosion remaining life of underground pipelines with a mechanically-based probabilistic model[J]. Journal of Petroleum Science and Engineering 65 (2009) pp 162-166.
- [7] Hyun-Young Chang, Yong-Soo Park, Woon-Suk Hwang. Initiation modeling of crevice corrosion in 316L stainless steels[J]. Journal of Materials Processing Technology 103 (2000) pp 206-217.
- [8] Fan Q.S., Yin Y.J., Yu W.J. (2013). *mechanics of materials*[M]. Tsinghua University Press.

- [9] A.K. Sheikh, J.K. Boah, D.A. Hanen. Statistical Modeling of Pitting Corrosion and Pipeline Reliability[J]. Corrosion Vol. 46, No. 3 (1990) pp190-197.

Condition-based maintenance for degrading systems with state-dependent operating cost

B. Liu*, M. Xie* and W. Kuo*

* Department of Systems Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong + email address: binliu9-c@my.cityu.edu.hk; minxie@cityu.edu.hk; way@cityu.edu.hk

Abstract. This paper develops a condition-based maintenance policy for degrading systems with consideration of the state-dependent operating cost. The state-dependent operating cost occurs when the system is in a deteriorating state, even if the system is still functioning. The system is subject to a continuous-time degradation process. Wiener process is used to characterize the evolution of system state. By taking advantage of the independent increment of Wiener process, we model the CBM policy as a Markov decision process. The system is subject to periodic inspection, at which decision has to be made whether to replace the system or let it be. The system is replaced when its degradation level exceeds the replacement threshold. We show that the optimal maintenance strategy is a monotone control limit policy. Compared with the CBM policy where no state-dependent operating cost is considered, our policy shows a more conservative replacement threshold.

1. Introduction

With the fast development of sensing techniques, system state can now be monitored or inspected at a much lower cost, which facilitates to characterize the deteriorating process by a continuous-time stochastic model. For systems subject to continuous-time degradation process, condition-based maintenance (CBM) has shown its effectiveness and predominance in preventing unexpected failures (Caballé et al, 2015). As the CBM utilizes the real-time information of the system health, it is usually more effective in reducing operating cost than the traditional age-based and block-based maintenance policies (Wu & Ryan, 2010; Liu et al, 2015). The ultimate goal of CBM is to reduce the risk of catastrophic failure and meanwhile reduce operating costs by eliminating unnecessary maintenance actions.

Concerned with CBM, a lot of research effort has been devoted to the prognostic models, with respect to characterizing the physical deterioration (Flory et al, 2015). The degradation process can be classified according to the degradation states: discrete or continuous (Liu et al, 2014). For system with discrete degradation states, Markov chain models are suitable to describe the deterioration process. The Markov chain models are useful when the degradation states cannot be accurately detected. However, it suffers the disadvantage of arbitrary classification of the degradation states.

Recently, thanks to the advances of sensing techniques, the system states can be monitored accurately, which facilitates the application of continuous degradation models in characterizing the physical deterioration (Chen et al, 2015; Elwany et al, 2011). Stochastic processes such as Wiener process, Gamma process and inverse Gaussian process, are widely used due to their feasible mathematical properties and clear physical interpretations. The property of independent degradation increments makes these stochastic models extremely attractive.

In traditional CBM models, optimal decisions are usually determined by minimizing the long-run cost rate for an infinite or a finite horizon. The cost items considered include the inspection, repair and replacement cost, and cost incurred by unexpected failures. It is assumed that during the operation period, no additional cost is charged as long as the system stays in the functional state, regardless of the degradation level. However, in reality, as the system degrades, its performance decreases as well, which results in increasing operating cost. For example, in a production system, if a machine deteriorates over time, the quality of the produced products would decrease accordingly, occurring additional cost (Xu and Cao, 2015). Therefore, it is more reasonable to take in account the state-dependent operating cost when making maintenance decisions.

This paper develops a condition-based maintenance policy for degrading systems with state-dependent operating cost. The state-dependent operating cost occurs due to system deterioration. The system is subject to a continuous-time degradation process, characterized as Wiener process. The CBM policy is modeled by

taking advantage of the independent increment of Wiener process. The system is subject to periodic inspection, at which decision has to be made whether to replace the system or let it be. The system is replaced at failure or when its degradation level hits the replacement threshold at inspection. We show that the optimal maintenance strategy is a monotone control limit policy.

The rest of this paper is organized as follows. Section 2 presents the maintenance model, where the degradation process is introduced and the maintenance framework is formulated. Section 3 explores the structural property of the maintenance policy. Numerical example is given in Section 4 to illustrate the monotone control limit policy. Finally, concluding remarks are given in Section 5.

2. Maintenance model

In this section, we present the degradation process of the system and formulate the maintenance cost function. We use Wiener process to describe the degradation process due to its mathematical properties and easy implement in practice. Maintenance cost includes not only the traditional cost items, but also the operating cost due to system degradation.

2.1 Description of degradation process

The system is subject to continuous degradation process, described as Wiener process. The general form of Wiener process is presented as

$$dX(t) = \mu(t)dt + \sigma(t)dB(t). \quad (1)$$

Taking integral of Equation 1, we have

$$X(t) = x_0 + \int_0^t \mu(s)ds + \int_0^t \sigma(s)dB(s) = M(t) + \Lambda(t) + x_0, \quad (2)$$

where $M(t) = \int_0^t \mu(s)ds$ and $\Lambda(t) = \int_0^t \sigma(s)dB(s)$.

Usually, the initial degradation level at installation is 0, $x_0 = 0$. Further, if the system is subject to a stationary Wiener process, which implies that μ and σ are constant, we can rewrite the degradation process as

$$X(t) = x_0 + \mu t + \sigma B(t). \quad (3)$$

For notational simplicity, we focus on stationary Wiener process in this paper, $\mu(t) = \mu$ and $\sigma(t) = \sigma$.

2.2 Maintenance formulation

We consider the CBM policy where periodic inspection is dedicated to monitor the degradation level of the system. Throughout the paper, we assume that the maintenance actions are perfect. As a result, the word “replace” and “maintain” are used interchangeably. Sudden failure occurs when the degradation level exceeds the failure threshold l . Correspondingly, corrective maintenance is carried out immediately to replace the failed system, with the corrective maintenance cost c_r . During the operation horizon, inspection is implemented periodically. At each inspection time, if the degradation level does not exceed the failure threshold, decision has to be made whether to preventively replace the system or let it be until the next inspection. The decision is made based on the current degradation level at inspection. If the system is replaced preventively, a preventive maintenance cost c_p is incurred. If we decide to wait until the next inspection, we have to pay the inspection cost c_i and the system runs the risk of sudden failure (during the inspection time and the next inspection). It is assumed that both preventive maintenance and corrective maintenance can restore the system to the as-good-as-new state. Apparently, $c_r > c_p$, because c_r includes additional cost due to sudden failures, such as production loss and even life hazards. For example, in the processing industry, an unexpected failure can lead to a stopping in processing line and even disasters such as fire and explosion, resulting in significant economic losses.

Apart from the aforementioned cost items, what cannot be neglected is the operating cost during system operation. This cost item occurs due to system degradation, which is used to represent the efficiency or quality loss. Generally, the operating cost not only relates to system degradation level, but also to system age. This is due to the fact that a more aged system is more likely to result in efficiency loss and quality reduction.

The system is inspected at equally spaced discrete time epochs, $\{\tau, 2\tau, 3\tau, \dots\}$, where τ is the time unit between two consecutive inspections. At the k th inspection epoch, denote system state as X_k and system age as $\tau_k = k\tau$, which is the time since the previous maintenance action. At each inspection when the system age and degradation level is detected, $\{k, X_k\}$ then constitutes the state of sequential decision process. We follow the assumption of Elwany et al (2011) that if the degradation level exceeds the failure threshold between inspections and then returns below the threshold at the next inspection, this does not incur a failure.

Assuming that the operating cost occurs when the system age is larger than $k_c\tau$ (k_c is a known integer), the operating cost of the system at age t and degradation level x is

$$G(t, x) = \begin{cases} 0, & t < k_c\tau \\ \alpha e^{\beta(t-k_c\tau)} x, & t \geq k_c\tau \end{cases}, \quad (4)$$

where α and β are known constant, scaling the effect of age and degradation level on the operating cost.

The incurred cost has a discounting factor e^{-rt} at time t , where r is a known constant, $r < \beta$. According to the model assumptions, the cumulative operating cost between τ_k and τ_{k+1} is denoted as

$$Q(k, X_k) = \int_{\tau_k}^{\tau_{k+1}} G(t, x) dt = \int_{\tau_k}^{\tau_{k+1}} \alpha e^{(\beta-r)(t-k_c\tau)} x(t) dt, \quad \tau_k > \tau_c. \quad (5)$$

The expected operating cost can be expressed as

$$\begin{aligned} W(k, X_k) &= E[Q(k, X_k)] = E\left[\int_{\tau_k}^{\tau_{k+1}} \alpha e^{(\beta-r)(t-k_c\tau)} x(t) dt\right] \\ &= E\left[\int_{\tau_k}^{\tau_{k+1}} \alpha e^{(\beta-r)(t-k_c\tau)} (X_k + \mu(t-\tau_k) + \sigma B(t-\tau_k)) dt\right] \quad k \geq k_c. \quad (6) \\ &= (X_k + \mu(k-k_c)) \frac{\alpha}{(\beta-r)} e^{(\beta-r)(k-k_c+1)\tau} - (X_k + \mu(k-k_c) - \mu) \frac{\alpha}{(\beta-r)} e^{(\beta-r)(k-k_c)\tau} \end{aligned}$$

Denote $V(k, X_k)$ as the infinite-horizon minimal total discounted cost with the initial state (k, X_k) . The optimality equation satisfies the Bellman equation, expressed as

$$V(k, X_k) = \begin{cases} c_f + V(0, 0) & X_k > l \\ \min \{c_p + V(0, 0), \\ e^{-r\tau} U(k, X_k) + e^{-r\tau} W(k, X_k)\} & X_k \leq l \end{cases}, \quad (7)$$

where $U(k, X_k) = E[V(k+1, X_{k+1}) | X_k]$ is the expected cost-to-go at the $(k+1)$ th inspection. Note that the inspection cost is treated as a separate cost item and not incorporated in Equation 7. The optimality equation follows the logic that (1) if the degradation level at the k th inspection exceeds the failure threshold l , which implies that the system has already been failed, the system has to be replaced correctively; (2) if the system still functions, we may choose either to preventively replace the system or let it be till the next inspection, depending on which is more cost-effective.

3. Structural property of maintenance policy

In this section, we focus on the structural properties of the maintenance policy, as structural properties are usually essential to develop computationally efficient algorithms. In particular, we concentrate on the control limit policy. Control limit policy implies that the system is kept operating until the observed degradation level exceeds a specific threshold. We present the essential properties as follows.

Theorem 1. The value function, $V(k, X_k)$, is non-decreasing in the inspection time k and non-decreasing in X_k .

Proof. We prove the theorem by mathematical induction. Denote $V^n(k, X_k)$ as the value function at the n th iteration of value iteration policy. At $n=0$, we have $V^0(k, X_k)=0$, which is a constant. Assume that the theorem holds for the n th iteration, i.e., $V^n(k, X_k)$ is non-decreasing in k and non-decreasing in X_k . Then according to Equation 7, we have

$$V^{n+1}(k, X_k) = \begin{cases} c_f + V^n(0, 0) & X_k > l \\ \min \left\{ c_p + V^n(0, 0), \right. \\ \left. e^{-r\tau} U^n(k, X_k) + e^{-r\tau} W(k, X_k) \right\} & X_k \leq l \end{cases} \quad (8)$$

Equation 6 clearly shows that $W(k, X_k)$ is non-decreasing in k and X_k . As $V^n(k, X_k)$ is non-decreasing in k and non-decreasing in X_k , its expectation, $U^n(k, X_k)$, holds this property as well. Because the terms of the right-hand side are non-decreasing in k and non-decreasing in X_k , $V^{n+1}(k, X_k)$ is also non-decreasing in k and X_k , which completes the proof. $\square$

Theorem 2. The optimal maintenance policy that minimizes the value function is a monotonic control limit policy. At decision epoch k , preventive replacement is performed when the observed degradation level exceeds the optimal value δ_k . The sequence $\{\delta_k, k=1, 2, \dots\}$ is non-increasing in k .

Proof. Preventive maintenance is optimal when $c_p + V(0, 0) < e^{-r\tau} U(k, X_k) + W(k, X_k)$. The left-side is a constant. Based on Theorem 1, $e^{-r\tau} U(k, X_k) + W(k, X_k)$ is non-decreasing in X_k . Thus for any $X_k > \delta_k$, the inequality holds, which implies the monotonic control limit policy. On the other hand, the term on the right-hand side is non-decreasing in k . Therefore, for the inspection epoch k , there exists an optimal replacement age k_δ , such that for any $k > k_\delta$, the optimal decision is to preventive replace the system, which indicates that δ_k is non-increasing in k . $\square$

Theorem 2 implies a non-increasing control limit policy, which is different from the work of Elwany et al. (2011) and Chen et al. (2015). This is due to the fact that in the work of Elwany et al. (2011) and Chen et al. (2015), there is a predicting step to estimate the degradation parameters. The predicting accuracy increases with k , which results in a more tolerance of older systems. On the other hand, in our paper, the operating cost shows an increasing trend with system age, which implies an urgency to replace the system with larger age.

Remark. If the operating cost is not incorporated into the cost function, then the maintenance policy is reduced to a constant control limit policy (Chen et al., 2015). This is due to the independent increment property of the Wiener process.

4. Numerical example

The degradation process of the system follows Wiener process with the drift coefficient $\mu=1$ and diffusion coefficient $\sigma=1$. The system fails when its degradation level hit the failure threshold $l=10$. It is assumed that the system is subject to periodic inspection, with the inspection interval $\tau=1$ and the inspection cost $c_i=0.05$. The incurred cost is discounted with the discounting factor $r=0.02$. At the k th inspection epoch, if the system functions but its degradation level exceeds the control limit δ_k , preventive replacement is carried out with cost $c_p=4$. If the system has already failed, then corrective replacement is performed with cost $c_r=10$. In the operation horizon, operating cost is incurred due to system degradation. The operating cost occurs when the inspection number is larger than $k_c=4$. The associated scaling parameters are $\alpha=0.2$ and $\beta=0.05$.

Given the monotonic property of the maintenance policy, the optimal policy can be obtained by many of the existing method, e.g., value iteration algorithm and policy iteration algorithm (Puterman, 2009). To make the solution feasible, we follow the existing practice to discretize the continuous state into finite intervals and use a time horizon long enough to approximate the infinite horizon.

The survival probability of the degradation process is shown in Figure 1. As can be observed from Figure 1, when the system ages over 10, the system will fail with the probability of 0.9. In addition, the system will fail with the probability of 0.99 after age 15.

Figure 2 shows how the optimal preventive maintenance threshold varies with system age. As the expected operating cost increases with the degradation level and system age, the optimal preventive maintenance threshold decreases with system age. To investigate the effectiveness of the monotone control limit policy, we compare the monotone control limit policy with the traditional constant control limit policy. When the control limit is set as 3.85, the total discounted cost reaches its minimum 52.62. By comparison, the minimum total discount cost under monotone control limit policy is 48.45. In addition, Figure 2 shows a constant control limit policy for system between age 0 and 4, where no operating cost is incorporated. Compared with the CBM policy where no state-dependent operating cost is considered, the proposed maintenance policy shows a more conservative replacement threshold.

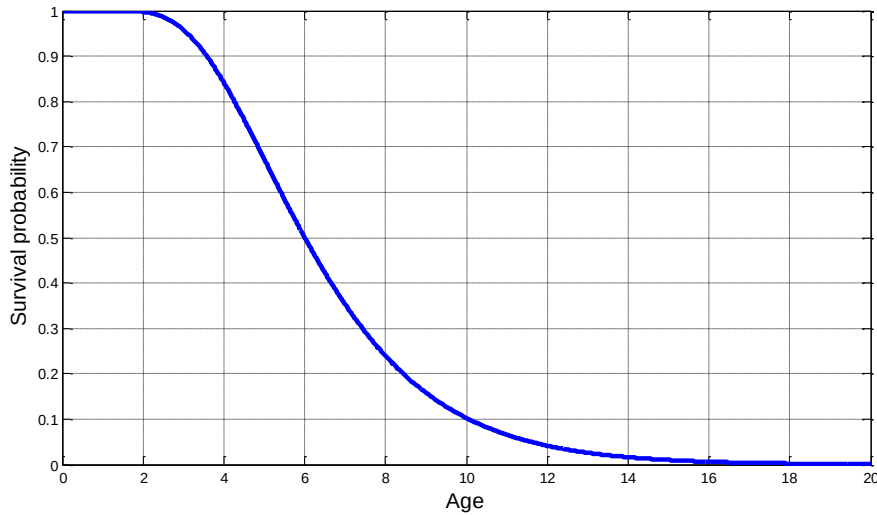


Figure 1. Survival probability of degradation process

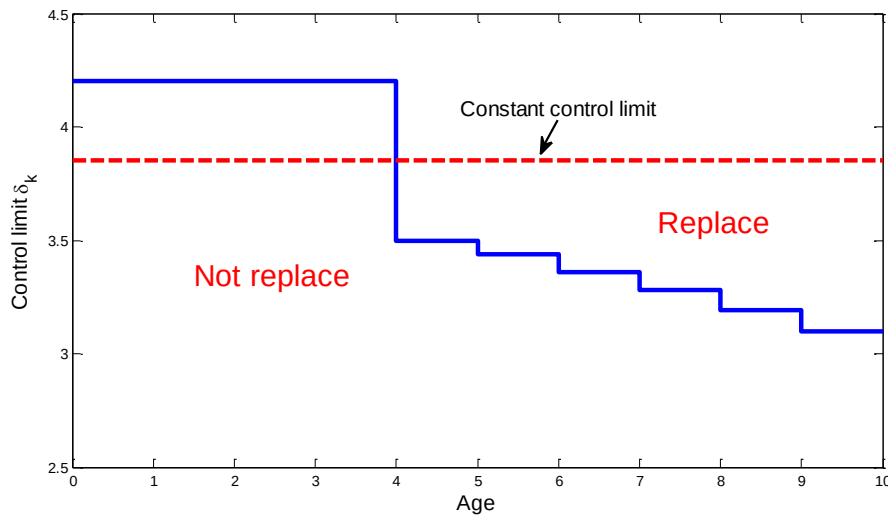


Figure 2. Optimal maintenance policy with system age

5. Conclusions

This paper investigates CBM policy for degrading systems with state-dependent operating cost. Evolution of the system state is described by Wiener process. We show that the optimal maintenance policy is a monotone control limit policy. Optimal preventive maintenance threshold is found to be non-increasing with system age, which is illustrated through a numerical example.

Further research can be conducted by generalizing the degradation process to a non-stationary stochastic process. In addition, this paper assumes that the system states can be accurately revealed at inspection, which can be extended to partial observable states.

References

- Caballé NC, Castro IT, Pérez CJ, Lanza-Gutiérrez JM (2015) A condition-based maintenance of a dependent degradation-threshold-shock model in a system with multiple degradation processes. *Reliability Engineering & System Safety* **134**, 98-109.
- Chen N, Ye ZS, Xiang Y, Zhang L (2015) Condition-based maintenance using the inverse Gaussian degradation model. *European Journal of Operational Research* **243**, 190-199.
- Elwany AH, Gebraeel NZ, Maillart LM (2011) Structured replacement policies for components with complex degradation processes and dedicated sensors. *Operations research* **59**, 684-695.
- Flory JA, Kharoufeh JP, Abdul - Malak DT (2015) Optimal replacement of continuously degrading systems in partially observed environments. *Naval Research Logistics* **62**, 395-415.
- Liu B, Wu J, Xie M (2015) Cost analysis for multi-component system with failure interaction under renewing free-replacement warranty. *European Journal of Operational Research* **243**, 874-882.
- Liu B, Xu Z, Xie M, Kuo W (2014) A value-based preventive maintenance policy for multi-component system with continuously degrading components. *Reliability Engineering & System Safety* **132**, 83-89.
- Puterman ML (2009) *Markov decision processes: Discrete stochastic dynamic programming*. Hoboken, NJ: Wiley-Interscience.
- Wu X, Ryan SM (2010) Value of condition monitoring for optimal replacement in the proportional hazards model with continuous degradation. *IIE Transactions* **42**, 553-563.
- Xu W, Cao L (2015) Optimal tool replacement with product quality deterioration and random tool failure. *International Journal of Production Research* **53**, 1736-1745.

Using observers and Bond Graph in modelling and multiple fault Identification

B.Maaref*, H.Dhouibi*, Z.Simeu-Abazi **, , H.Messaoud*

* LARATSI, National Engineering School of Monastir

** G-SCOP, Grenoble- IPN/UJF- Grenoble 1, CNRS, France

Abstract. This article proposes a diagnostic approach for complex hybrid systems through the use of bond graph and observers well known in the continuous domain, coupled with timed automata used in the field of discrete event systems. Inspired by the signature analysis of residues (using observer model), the sensor signatures (using timed automata model), a global model based on Bond Graph, the observer and the automata is studied for solve the problem of identifying multiple faults where the system suffers from several faults (the actuator faults and sensor fault). The proposed approach is validated by simulation tests in a hydraulic system with two tanks.

1. Introduction

The increasing complexity of industrial systems which have become increasingly demanding in terms of reliability, performance, security and availability constraint, has given increasing interest in diagnosis. This last fits into the overall framework of the monitoring and supervision. Therefore, for a great number of applications, it is necessary to implement a surveillance system to detect, isolate and identify any malfunctions (Cyrille Christophe 2001). The objective of this paper is the use of Bond Graph, well known in the continuous domain (Ould Bouamama B. et al. 2006), for modeling of physical processes (J.U. Thoma and B. Ould-Bouamama 2000, P. C. Breedveld 1986) on all domains (electrical, mechanical, thermal, ...) and the use of theory observers (P.M. Frank 1996, R.J. Patton 1994, J. Chen and R.J. Patton 1999), also well known in the continuous domain, for robust diagnosis (taking into account parameter uncertainties), coupled with timed automata (E. Gascard and Z. Simeu-Abazi 2013) used in the domain of discrete event systems that allow take into account the dynamic evolution system and their propagation failures (Bengtsson J. and Wang Yi 2004, Saeed Ghasemi et al. 2014). We have extended our approach to account for the diagnosis problem of multiple faults. It is multiple faults that have effects which overlap in time for diverse reasons such as the spread. We use detection and location techniques inspired by analysis residues matrices (from observer model) and sensor matrices (from the timed automata model).

This paper is organized as follows: Section 2 presents a proposed approach for the diagnosis of multiple faults. An illustrative example of a two-tank system is developed in section 3 and shows the effectiveness of the proposed method. At the end of a conclusion is presented with few prospects.

2. Approach proposed for the multiple faults diagnosis

The main performances sought in a diagnostic system, which is the ability of the latter to detect and isolate a fault from its appearance, corresponds the concept of detection and localization of multiple faults (V. Reppa et al. 2015, Imtiaz Fliss and Moncef Tagina 2009). The faults detection consists of analyzing residues (using the observer model). Thus, the multiple faults localization is generally analyzed on the basis of fault signatures matrices: residues matrices of observer model and sensors matrices of timed automaton model. The different steps to search for common faults 2-2 in general case are presented in the following chart, Figure 1.

3. Application to two tank system

3.1. Description of the system

The process, Figure 2, is composed of two cylindrical tanks: the first tank T1, section S1 of height h_1 and the second tank T2, section S2 and height h_2 . The tanks communicate by the intermediate a valve V1 (hydraulic resistance RV_1), always open. The process includes an only entry: volume flow q_i . The outlet flow of second tank is authorized by a valve V2 of hydraulic resistance RV_2 . We have put four sensors: two level sensors L1 and L2 and two overflow sensors L3 and L4.

Its global objective is to maintain a fluid level in the two tanks at a level well defined:

$$\begin{cases} h_1 \leq 0.15 \text{ m} \\ h_2 \leq 0.10 \text{ m} \end{cases} \quad \text{With } h_1 \geq 0.15 \text{ m} \Rightarrow L_1 = 1; h_2 \geq 0.10 \text{ m} \Rightarrow L_2 = 1$$

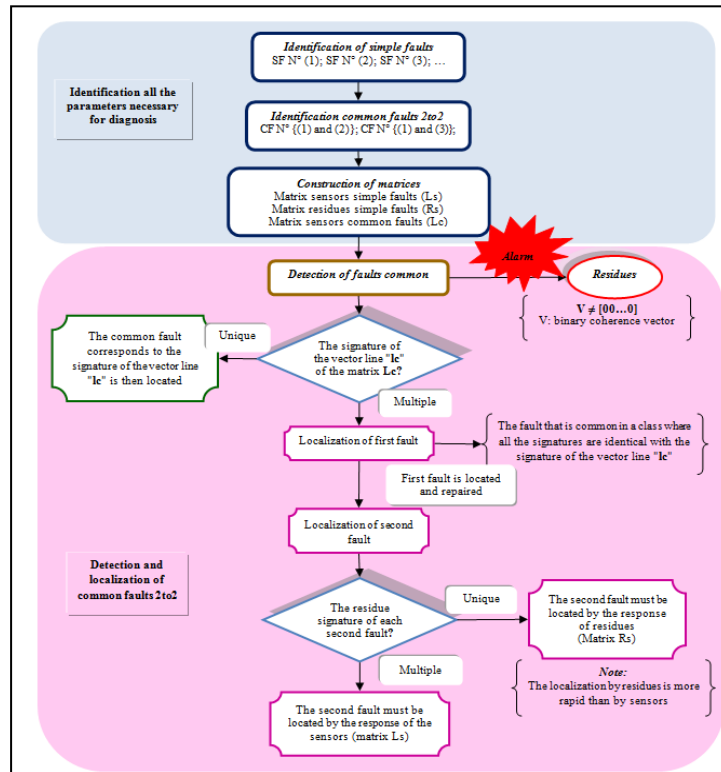


Figure 1. Flowchart of approach used for research common faults.

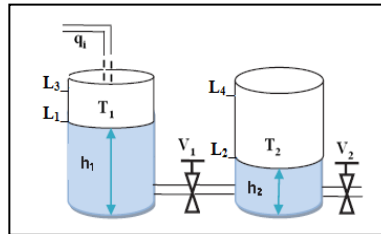


Figure 2. Two tanks hydraulic system.

3.2 Modelling by Bond Graph model

We have used a procedure described in (P. Borne et al. 1992) and (G. Dauphin-Tanguy 2000), the bond graph model of the physical system of Figure 2 is represented in Figure 3.

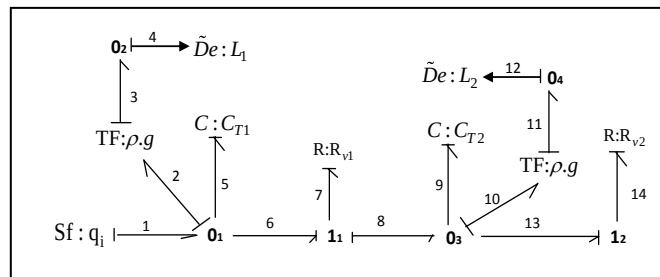


Figure 3. Bond Graph model of two tanks system

Using the Bond graph representation, system state equations can be directly deduced from the Bond graph model of the system. Then the state equation of a two-tank system (3), is obtained from (1) and (2) generated by the Bond graph model in Figure 3.

$$\begin{cases} S_1 \frac{dL_1}{dt} - q_i + \frac{\rho g(L_1 - L_2)}{R_{v1}} = 0 & (1) \\ S_2 \frac{dL_2}{dt} - \frac{\rho g(L_1 - L_2)}{R_{v1}} + \frac{\rho g L_2}{R_{v2}} = 0 & (2) \end{cases} \quad \begin{array}{c} \text{State} \\ \text{representation} \end{array} \quad \begin{cases} \frac{dh_1}{dt} = \frac{q_i}{S_1} - \frac{(h_1 - h_2)}{S_1 R_1}; \quad R_1 = \frac{R_{v1}}{\rho g} \text{ and } L_1 = h_1 \\ \frac{dh_2}{dt} = \frac{h_1 - h_2}{S_2 R_1} - \frac{h_2}{S_2 R_2}; \quad R_2 = \frac{R_{v2}}{\rho g} \text{ and } L_2 = h_2 \end{cases} \quad (3)$$

The system can be brought back in a stable closed loop system by the state feedback where K is the state feedback gain and v(t) is input for the closed loop system. The state equation obtained, (3), is thus in the form:

$$\begin{cases} \dot{x}(t) = A.x(t) + B.u(t) \\ \text{avec } u(t) = -K.x(t) + v(t) \\ y(t) = C.x(t) \end{cases} \quad (4)$$

With

$$A = \begin{pmatrix} \frac{-R_1}{S_1} & \frac{R_1}{S_1} \\ \frac{R_1}{S_2} & -\frac{(R_1 + R_2)}{S_2} \end{pmatrix}; B = \begin{pmatrix} \frac{1}{S_1} \\ 0 \end{pmatrix}; C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; K = [-4.7 \ 6.06]$$

3.3 Diagnosis model observer-based

The method by observers consists to reconstruct, from the mathematical model, (4), the outputs system through an estimate the system states.

$$\text{system : } \begin{cases} \dot{x}(t) = A.x(t) + B.u(t) \\ y(t) = C.x(t) \end{cases} \quad \text{Observers : } \begin{cases} \dot{\tilde{x}}(t) = A.\tilde{x}(t) + B.u(t) + L(y - \tilde{y}) \\ \text{avec } u(t) = -K.\tilde{x}(t) + v(t) \\ \tilde{y}(t) = C.\tilde{x}(t) \end{cases}$$

$$A = \begin{pmatrix} \frac{-1}{S_1.R_1} & \frac{1}{S_1.R_1} \\ \frac{1}{S_2.R_1} & -\frac{1}{S_2} \left(\frac{R_1 + R_2}{R_1.R_2} \right) \end{pmatrix}; B = \begin{pmatrix} \frac{1}{S_1} \\ 0 \end{pmatrix}; C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; K = [-4.7 \ 6.06]; L = \begin{pmatrix} -120 & 130 \\ 130 & -175 \end{pmatrix}$$

And this method capable of solving the fault detection and location problem by the evolution of residues : $r(t) = \varepsilon(t) = y(t) - \tilde{y}(t)$.

3.4 Matrix residues simple faults (Rs)

The residuals generation consists in comparing measurements stemming from the system to their estimations stemming from observer-based model. Only one residual allows the detection of fault. However, the fault location requires a set of structured residuals. These residues led to the matrix residues simple faults (Rs) which is given by Table 1.

Table 1. Matrix residues simple faults (Rs)

Components	R ₁	R ₂
Valve V ₁	1	0
Valve V ₂	1	1
Level sensor L ₁	1	0
Level sensor L ₂	0	1

3.5 Diagnosis model based on timed Automata

A timed automata is a finite state machine provided with an assembly continuous variables in pieces called clocks. These variables measure the passage of time. Our objective, through the use of timed

automata, is to build a dynamic model that contains all possible states (normal and failed states) system, which allows us to follow its temporal evolution.

Inspired by the analysis of faults signatures, we are going envisage a multiple fault scenario, which means that common faults can happen on the system.

3.6 Fault signature analysis

In order to isolate a fault, the method described in this article uses the default signature notion. A failure can be characterized by its signature, ie a set of values that take 0 and 1 for precise failure. It can be represented as a diagnostic matrix using the timed automata model. The matrix of sensors simple defects (Ls), Table 2, is constructed in the following manner:

- The lines are each failure mode (single fault)
- The columns represent each value that can take the variable sensor for each single fault.

Table 2. Matrix of sensors simple faults (Ls)

N° SF*	Failure mode	Sensors			
		L ₁	L ₂	L ₃	L ₄
(1)	Valve V ₁ Stuck_Close	1	0	1	0
(2)	Valve V ₂ Stuck_Close	1	1	1	1
(3)	Valve V ₂ Stuck_Open	0	0	0	0
(4)	Level sensor L ₁ Stuck_Close	1	0	0	0
(5)	Level sensor L ₁ Stuck_Open	0	1	1	1
(6)	Level sensor L ₂ Stuck_Close	0	1	0	0
(7)	Level sensor L ₂ Stuck_Open	1	0	1	1

* SF: Simple fault

And in the same manner, the matrix of sensor common faults (Lc) is constructed to take into account that the defects are multiple (common faults 2-2), Table 3.

Table 3. Matrix of sensors common faults (Lc)

N° CF*	Sensors			
	L ₁	L ₂	L ₃	L ₄
(1) and (2)	1	0	1	0
(1) and (3)	1	0	1	0
(1) and (4)	1	0	1	0
(1) and (5)	0	0	1	0
(1) and (6)	1	1	1	0
(1) and (7)	1	0	1	0
(2) and (4)	1	1	1	1
(2) and (5)	0	1	1	1
(2) and (6)	1	1	1	1
(2) and (7)	1	0	1	1
(3) and (4)	1	0	0	0
(3) and (5)	0	0	0	0
(3) and (6)	0	1	0	0
(3) and (7)	0	0	0	0
(4) and (6)	1	1	0	0
(4) and (7)	1	0	1	1
(5) and (6)	0	1	1	1
(5) and (7)	0	0	1	1

* CF: Common fault.

4. Simulation results

The diagnostic method proposed in this article has been validated. Indeed, for each injection of multiple faults random manner, the diagnostic approach is able to detect and isolate multiple faults at any instant and effective manner.

We injected a common fault (CF No. (2) and (6)) faults on the valve V2 (Valve V2 Stuck_Close, fault No. (2)) and faults on the sensor L2 (Level sensor L2 Stuck_Close fault No. (6)) at instant $t = 3s$.

The residues answers to this failure are given in Figure 4. The instant when the residues R1 and R2 exceed their respective thresholds ($R1 \neq 0$ and $R2 \neq 0$) represents the instant of detection of common fault. $T_{\text{detection}} = 3s$.

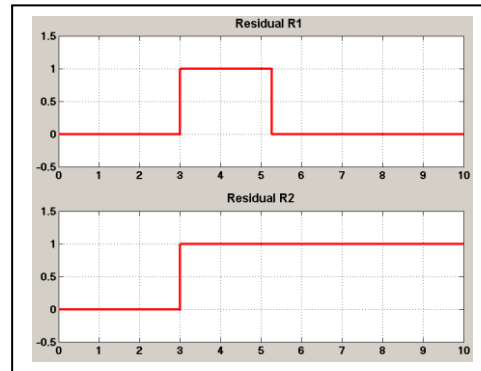


Figure 4. Residues response in the presence of a common fault No. (2) and (6)

Figure 5 show the state of sensors L1, L2, L3 and L4 and their sensitivity to this common fault. If we refer to the common fault signature No (2) and (6) given in Table 3, we find that this signature is confused with the signature of the common fault No (2) and (4) (the blue lines of Table 3) and therefore this common fault is only detectable but not localizable.

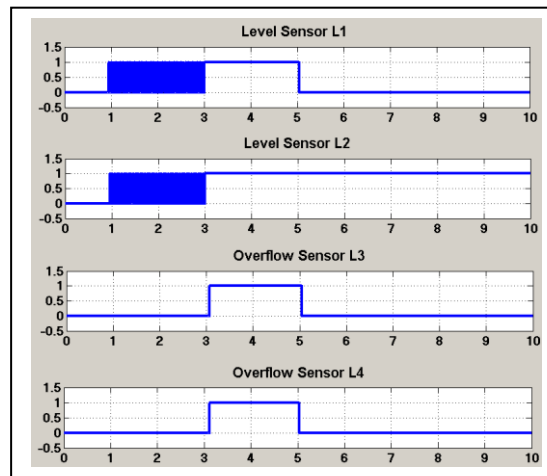


Figure 5. The state of sensors in the presence of a common fault No. (2) and (6)

To resolve this problem, we can locate, first of all, the fault that is common in a class (blue lines of Table 3) where all the signatures are identical with the common fault signature No (2) and (6), this is the fault No. (2), table 3, $T_{\text{location 1st fault}} = 3,11s$.

As soon as the fault is located and repaired, the instant when $T = 5s$, the residues answer change, Figure 4, also that for the state of the sensors, Figure 5, and these new responses allow to identify the second fault. If we refer to the second fault signature (fault level sensor L2 equivalent to fault No. (6)) given in Table 1, we find that this signature is unique and this result is well in conformity with what is expected; in other words, in the case of the second fault (6) only the residue R2 exceed their threshold. This instant represents the instant of location of second fault. $T_{\text{location of the 2nd fault}} = 5,26s$.

5. Conclusion & perspectives

With respect to existing works in this article, our contribution is based on the diagnosis of multiple faults using Bond Graph, the Observer and the Timed Automata. The proposed approach allows to solve the problem of identification multiple faults. Indeed, for a complex system, the Bond Graph allows the modelling of multi-energy physical processes (Electrical, mechanical, thermal ...). The approach based on the observer is used to generate indicators of faults (or residues). The fault detection algorithm consists to analyze these residues. Thus, the multiple fault isolation procedure consists in analyzing the faults of signatures matrices: matrices of residues from the observer model and matrices of sensors from timed automaton model. The timed automata allows to take into account the dynamic evolution of the system and fault propagation. The proposed approach is validated by simulation tests in a hydraulic system with two tanks.

It is expected that the results obtained can equally be extended to the location of multiple faults in real systems. In fact, we intend in future works for highlighting the possibility of using such a method in real application.

References

- Bengtsson J. and Wang Yi (2004) *Timed automata: semantics, algorithms and tools*. In lecture Notes on Concurrency and Petri Nets, W. Reisig and G. Rozenberg (eds), LNCS 3098, Springer-Verlag.
- Borne P., Dauphin-Tanguy G., Richard J. P., Rotella F., Zambettakis I. (1992) *Modélisation et identification des processus*. Editions Technip.
- Breedveld P. C. (1986) *A systematic method to derive bond graph models*. Int Journal of modeling and simulation, Vol. 1, No. 1, pp.57-62.
- Chen J. and Patton R.J. (1999) *Robust model-based fault diagnosis for dynamic systems*. Kluwer academic publishers.
- Cyrille Christophe (2001) *Surveillance des Systèmes Non Linéaires, Application aux machines électriques*. thèse de doctorat, Laboratoire d'Automatique et d'Informatique de Lille, UPRES A 80 21.
- Dauphin-Tanguy G. (2000) *Les bond graph*. HERMES Sciences Publications.
- Frank P.M.(1996) *Analytical and qualitative model-based fault diagnosis - a survey and some new results*. European Journal of Control, vol. 2, pp. 6–28.
- Gascard E., Simeu-Abazi Z. (2013) *Modular Modelling for the Diagnostic of Complex Discrete-Event Systems*. IEEE Transactions on Automation Science and Engineering IEEE Trans. Autom. Sci. Eng. (T-ASE-2012-244), vol.10, Issue 4, pp 1101-1123.
- Imtiaz Fliss and Moncef Tagina. (2009) *Multiple faults diagnosis using causal graph*. In the proceedings of The 6th IEEE International Multi- Conference On Systems, Signals Devices- SSD'09, Djerba, Tunisia, 23-26 Mars.
- OuldBouamama B., Medjaher K., Samantaray A.K., Staroswiecki M. (2006) *Supervision of an industrial steamgenerator. Part I: Bond graph modeling*. Control Engineering Practice, vol. 14, pp. 71–83.
- Patton R.J. (1994) *Robust model-based fault diagnosis: the state of the art*. In IFAC SafeProcess Symposium, Espoo, Finland, pp. 1–24.
- Reppa V., Marios M., Christos G. (2015) *Decentralized Isolation of Multiple Sensor Faults in Large-Scale Interconnected Nonlinear Systems*. IEEE Transactions on Automatic Control , Vol. 60 ,NO. 6.
- Saeed Ghasemi, Leili Mohammad Khanli, Mina Zolfi, GhaderTahmasebpour (2014) *Modeling and Simulation of NFC Logical Layer Peer-to-Peer Mode using CPN and TA*. International Journal of Electrical and Computer Engineering (IJECE), Vol. 4, No. 2, pp. 162~168.
- Thoma J.U., OuldBouamama B. (2000) *Modelling and simulation in thermal and chemical engineering: Bond graph approach*. Springer Verlag.

Agent-based Maintenance System Methodology for optimal Reconfigurable Vibrating Screen Management

O. A. Makinde*, K. Mpofu**, B. I. Ramatsetse*** and M. K. Adeyeri****

*Department of Industrial Engineering, Tshwane University of Technology, Pretoria, South Africa
olasumbomakinde@gmail.com

**Department of Industrial Engineering, Tshwane University of Technology, Pretoria, South Africa
mpofuk@tut.ac.za

***Department of Industrial Engineering, Tshwane University of Technology, Pretoria, South Africa
RamatsetseBI@tut.ac.za

****Department of Mechanical Engineering, Federal University of Technology, Akure, Nigeria
adeyerimichaeltut@gmail.com

Abstract. Reconfigurable Vibrating Screen (RVS) machine is an innovative beneficiation solution designed to ensure the screening of different mineral particle sizes and volumes demanded by the customers at any particular time in a cost-effective manner. In order to achieve RVS machine optimal functionality, reliability and maintainability, there is a need to develop an agent-based maintenance system, which is infused with the necessary knowledge base management and intelligent machine health monitoring systems such as diagnosing and prognosing algorithms/ tools that will be used for effectively maintaining and managing this machine when utilized in the mining industries. In view of this, this paper proposes a conceptual methodology required to develop an agent-based maintenance system that will ensure optimal RVS machine management. This agent-based methodology discusses the RVS machine diagnosing, prognosing and maintenance planning and inventory algorithms, required for developing its intelligent maintenance system. Successful application of this methodology will demonstrate to the mining community (RVS machine users), how RVS machine can be optimally maintained and managed to achieve its “near zero machine downtime” condition during its lifecycle.

1. Introduction

Fluctuations in mineral concentrates demand and production of varying sizes of mineral concentrates required by the customers had called for the design and development of a reconfigurable beneficiation solution that could be used to address the aforementioned problems. To this effect, an innovative beneficiation solution called Reconfigurable Vibrating Screen (RVS) was designed and developed in Tshwane University of Technology, South Africa to ensure dynamic and stochastic beneficiation processes required to produce mineral products demanded by the customers at any time “t”. The RVS machine is made up of different subsystems such as screen panel, rosta suspensions, side plates, vibrating motor, bearings, springs, torsion bar, side plates and back plates etc. These components all work together for its optimal functionality as depicted in Figure 1 (Ramatsetse, 2014). Due to the nature of the operations that the RVS machine is required to meet (i.e. the RVS machine is designed to screen different mineral particles such as coal, diamond, granite, limestone, iron and gold etc. of different specific gravities; screen different sizes and volumes of the aforementioned mineral particles demanded by the customers as well as mark-up production loss that might erupt during its repair and maintenance phase in order to meet the company’s monthly and annual production target), there is a need to develop an agent-based maintenance system which is infused with intelligent algorithms that is used for diagnosing (detecting) each of the subsystems of the RVS machine when it fails; prognosing (predicting) when each of the subsystems of the RVS machine will fail for aggregate maintenance planning of the RVS machine

and also forecasting the future mineral concentrates demand as yardstick for reconfiguration planning of the RVS machine; and ascertaining the inventory systems required in maintaining the RVS machine at different configurations. To develop this agent-based maintenance system, a holistic and comprehensive methodology, which highlights the chronological algorithms and modules needed to develop this system must be established. In view of this, the main aim of this paper is to develop a holistic methodology required to develop an agent-based system that is required to optimally manage and maintain the RVS machine. To achieve this, the first section of this paper reviews different maintenance systems that had been used to maintain machines in different manufacturing and mining industries while the second section presents and discusses the holistic methodology required to optimally maintain and manage the RVS machine when utilized in the mining industries.

2. Machine maintenance management systems

Machine maintenance is vital to ensure the upkeep of different subsystems that makes up a machine in order to ensure its functionality, usability, reliability and maintainability (Al-Najjar and Alsyouf, 2003; Lee, 2001; Dhilon, 2006 and Ashayeri, 2007). To this effect, different maintenance management systems such as corrective breakdown maintenance, which uses a fail and fix maintenance technique (Dhillon and Liu, 2006); preventive maintenance which uses “regular and routine observations” maintenance techniques (Sheu and Kuo, 2006, Mori and Fujishima, 2013); opportunistic maintenance that utilizes “repair opportunity” maintenance techniques (Cui and Li, 2006); predictive maintenance techniques, which utilizes probabilistic, stochastic and adaptive models for machine subsystems failure prediction (Shrotri and Khandagale, 2012); total productive maintenance which uses industrial tools and techniques such as Lean Six Sigma, FMEA and FMECA for machine maintenance optimization (Chan *et al.*, 2005 and Rajput and Jayaswal, 2012); dynamic maintenance which considers customers product demand in formulating machine maintenance policies (Adeyeri and Kareem, 2012); and computerized/e-maintenance systems which utilizes Information Communication and Technology (ICT) in synthesizing the different maintenance planning and scheduling, diagnosing and prognosing algorithms for optimal machine management have evolved over the years (Zhang *et al.*, 2003, Muller *et al.*, 2008, Holgado and Macchi, 2014). Based on the literature survey carried out on maintenance so far, it can be affirmed that an advanced maintenance system in form of an agent-based maintenance system (computer system) equipped with reactive, communicative, cognitive, autonomous, adaptive and rational intelligent behaviour (Maalal and Addou, 2011) have not been fully exploited and implemented in the manufacturing industries (Mourtzis *et al.*, 2016). To this effect, there is a need to formulate intelligent diagnosing, prognosing, maintenance planning and inventory algorithms (for Just-In-Time (JIT) maintenance of the different subsystems of the machine when it fails) required to ensure optimal functionality and reliability of the machine when used for industrial mining operations. These algorithms are orchestrated to be infused or integrated into RVS machine agent-based maintenance system. Agent system in the context of this work is an encapsulated computer system that is situated in some environments and capable of being flexible and autonomous in action in order to meet its design objectives (Telgen *et al.*, 2012). To achieve this, detailed chronological processes required to develop these aforementioned RVS maintenance management algorithm are discussed extensively in the next section of this paper.

3. Agent-based maintenance methodology for optimal RVS machine management

The agent-based maintenance methodology intends to discuss how intelligent algorithms required to optimally manage RVS machine will be designed and developed.

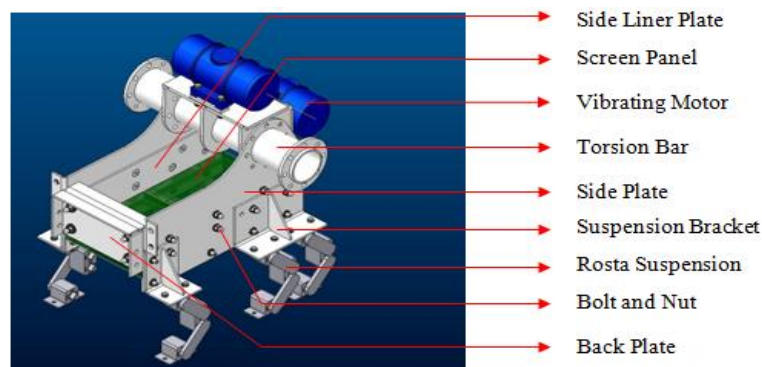


Figure 1: 3D representation of the RVS machine

This agent-based maintenance methodology when developed for the RVS machine management is expected to answer the following rhetorical questions: i) What is the reconfiguration size and time of the machine when customer demand changes? ii) If customer demand changes, what is the amount of RVS subsystems or spare part modules required to reconfigure the machine to meet the customer demands? iii) If the RVS machine breakdown for a particular of period time, by what size should the RVS machine be reconfigured in order to mark up its production loss? iv) When will the RVS machine subsystem(s) fail or degrade? v) At what exact time do subsystems of the RVS machine need to be repaired or replaced?

The agent-based maintenance system for optimal RVS machine is made up of three different intelligent agent systems called RVS machine diagnosing agent, RVS machine prognosing agent and RVS maintenance planning and inventory management agent as depicted in Figure 2. The RVS machine diagnosing agent of the RVS machine management system is equipped with diagnosing algorithms which answers the fifth rhetorical question; while the RVS machine prognosing agent of the RVS machine management system is equipped with prognosing algorithms which answers the fourth rhetorical question. Also, the RVS machine maintenance planning and inventory management agent is equipped with maintenance planning and inventory management algorithms which answer the first, second, third and fourth rhetorical questions. These aforementioned algorithms responsible for optimal functionality of these agents depicted in Figure 2 are discussed extensively in the next section of this paper. It is worthy to note that there is interoperability and interactions between these agents as clearly shown by arrows in Figure 2.

3.1 RVS machine diagnosing algorithms

The functional state of the different subsystems of the RVS machine is strictly monitored in real-time using intelligent sensors to determine when each of these subsystems will fail. The algorithms used in achieving this is stated as follows: i) Acquiring raw performance data of the different subsystems of the RVS machine in real time ii) Compare (i) above, with some rule-based algorithms (i.e. the maximum operating working conditions of the RVS machine system before it fails) that is obtained from the Finite Element Analysis (FEA) performance simulation of the different subsystems of the RVS machine using heuristic techniques in order to ascertain the time when the RVS subsystems will fail. iii) Initiate, monitor and measure the stress and deformation exhibited by the back liner plates, back plates, side plates, side liner plates, screen panel, screen panel pins, torsion bar and torsion bar bracket of the RVS machine using strain gauges; the tension and compression exhibited by the bolts holding different subsystems of the RVS machine using torque and tension sensors as well as the temperature, current and voltage utilized by the vibrating motor in powering the screening operation of the machine at different configurations of the machine using thermocouples,

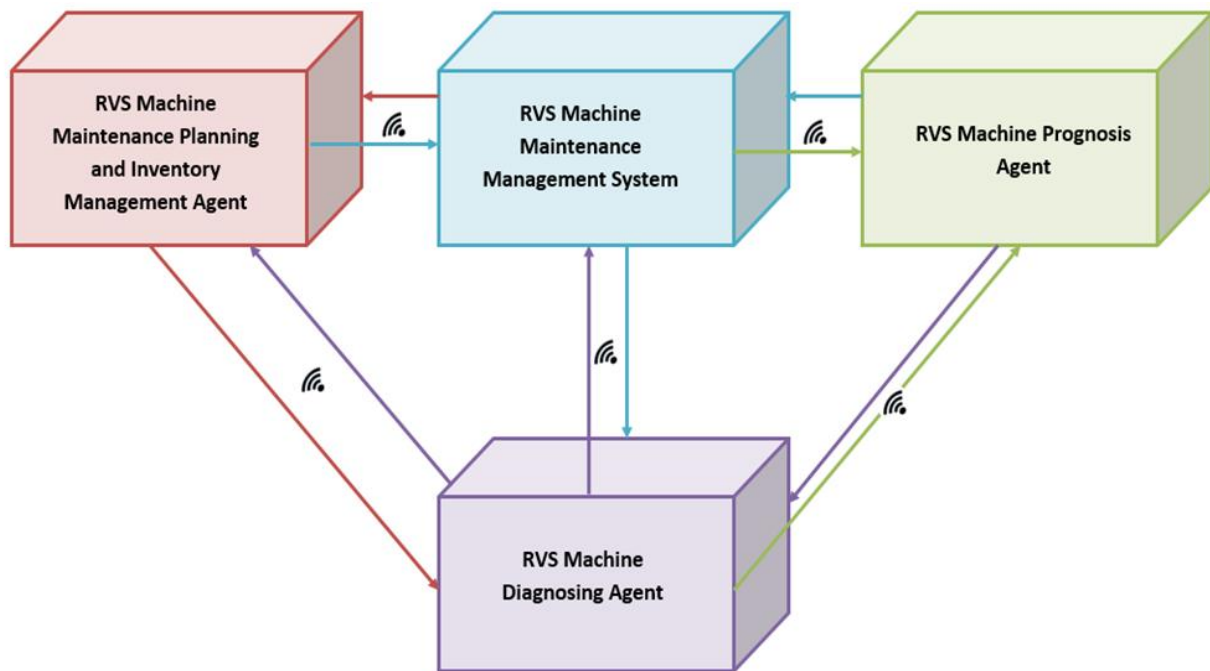


Figure 2: Agent-based Maintenance System Framework for optimal RVS Machine Management

current sensors and voltage sensors respectively. Note: an intelligent diagnosing system named, National Instrument Real Time (RT) CompatRio monitoring system will be used to achieve the aforementioned tasks. The system is made of intelligent data acquisition and analysing modules (such as thermocouple analysing modules, strain gauge analysing modules, accelerometer modules, current analysing modules, and voltage analysing modules e.t.c.), LabVIEW software and the host computer. iv) Processing and the interpretation of the acquired raw performance data of the RVS machine using Fast Fourier Algorithms (FFT). Please note that the LabVIEW software is used for coding the rule based algorithms that is used in making a decision (i.e. whether each subsystem is in good condition or not) in order to determine when to replace any subsystem of the RVS machine as well as developing a user interface that displays the current functional performance of different subsystems of the RVS machine. v) Lastly, display the real time performance result of the different subsystem of the RVS machine with a view to ascertain when each subsystem need to be repaired or replaced using the host computer. The front panel or user interface of a real time system used for monitoring the temperature condition of a RVS machine is depicted in Figure 3. This user interface displays the machine temperature and associated temperature warning strings and alarms, which informs whether the machine is running at normal or abnormal temperature. Furthermore, the codes which incrypt the rule based algorithms, used in determining the health (temperature) status condition of the RVS machine is synthesized in the LabView code block diagram depicted in Figure 4. The time and date at which the different subsystems of the RVS machine is malfunctioning (i.e. the failure time and date) is recorded in the LabView database tool kit. The rule-based algorithms used in making decision on when each subsystem of the RVS machine (configured at a size of 600mm × 1200mm for screening crushed granite of bulk density of 1554kg/m³) will fail (obtained from the RVS machine designer's manual) is highlighted in Table 1.

3.2 RVS prognosing algorithms

Intelligent offline prognosing algorithms, which models the date and time of failures obtained from the LabVIEW database tool kit of the diagnosing algorithms as well as historical mineral concentrates demand are developed to predict when the different subsystems of the RVS machine will fail as well as accurately forecast the future demand of mineral concentrates, which will serve as a yardstick to determine when to, by what amount and duration needed to reconfigure the RVS machine. The algorithms outlined for this are stated as follows: i) Invoke Holt's model adopted from Lepojec and Andelkovic-Pesic (2011) and Markov Model adopted from Cholette and Djurdjanovic (2014) which is meant for prognosis actions of the RVS subsystems failure and demand prediction due to their potentials in smoothing the errors and assigning weights to more recent data points during the prognosis process.



Figure 3: User Interface of the Host Computer of the Real-Time Monitoring System.

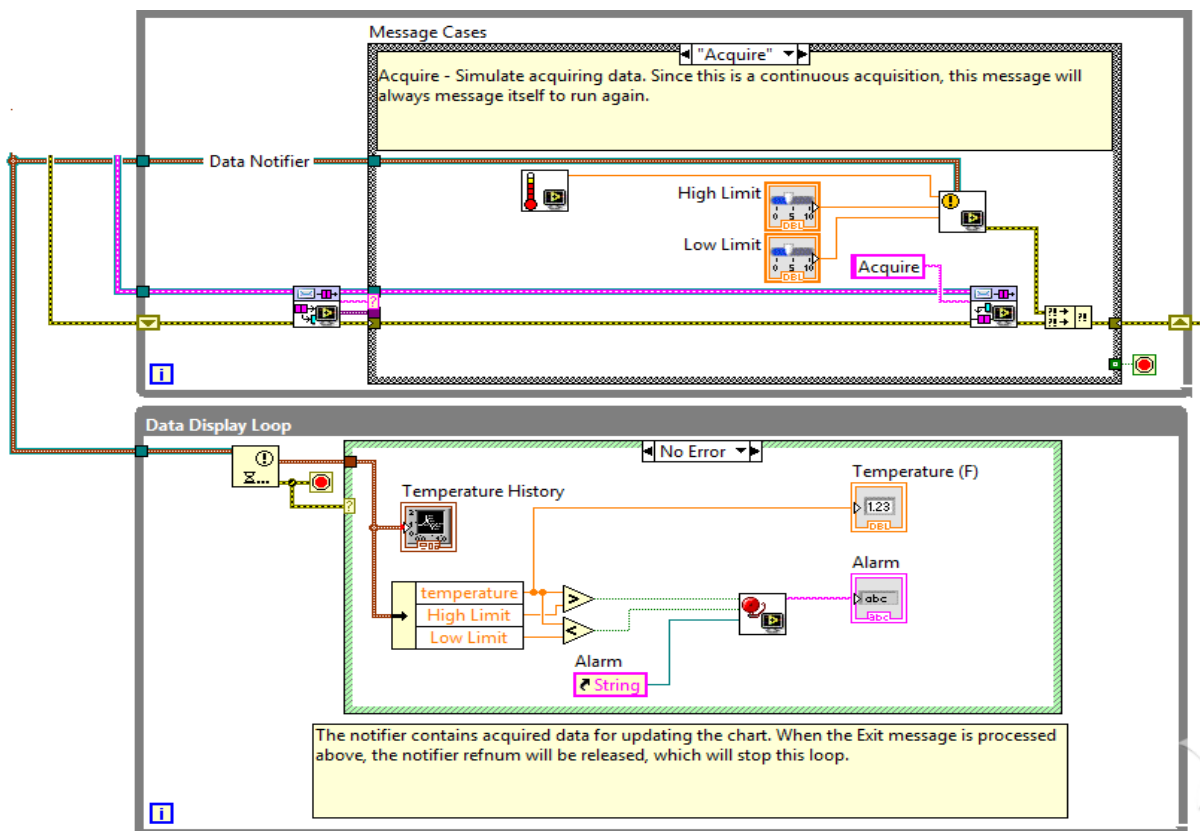


Figure 4: Block diagram for RVS Machine Temperature Measurement and Warning System

Table 1: Rule-based Algorithms for monitoring the performance behaviour of different subsystems of the RVS machine

Rule	Triggering Context
$\sigma_{torsion\ bar} < 9.95\text{MPa}$, $\sigma_{side\ plate} < 6.1\text{MPa}$, $\sigma_{rosta\ suspension} < 2.9\text{MPa}$, $\sigma_{back\ plate} < 14.2\text{MPa}$ and $\sigma_{screen\ panels} < 15\text{MPa}$	No warning – RVS subsystems are in healthy conditions.
$\sigma_{torsion\ bar} \geq 9.95\text{MPa}$, $\sigma_{side\ plate} \geq 6.1\text{MPa}$, $\sigma_{rosta\ suspension} \geq 2.9\text{MPa}$, $\sigma_{back\ plate} \geq 14.2\text{MPa}$ and $\sigma_{screen\ panels} \geq 15\text{MPa}$	Heat stroke warning - RVS subsystems are in faulty conditions.
Vibration speed of the RVS machine $< 0.283\text{m/s}$	Freeze Warning – RVS machine is under-functioning.
Vibration speed of the RVS machine $= 0.283\text{m/s}$	No Warning – RVS Machine is functioning at normal operating conditions.
Vibration speed of the RVS machine $> 0.283\text{m/s}$	Heat Stroke Warning – RVS Machine is functioning abnormally.
Outward temperature of the RVS machine $\leq 80^\circ\text{C}$	No Warning - RVS Machine is functioning at normal operating conditions.
Outward temperature of the RVS machine $\geq 80^\circ\text{C}$	Heat Stroke Warning – RVS Machine is functioning abnormally.

Where, σ denotes the stress.

For explicit explanation, the detailed prognosing algorithm used in determining when to replace or repair the different subsystems of the RVS machine; forecasting the future mineral concentrates demand and also predicting the remaining effective useful life of the RVS machine after its usage for a particular period of time ‘t’ are illustrated in equation 1 to 3.

$$F_{t+1} = a_t + b_t \quad (1)$$

Where, $a_t = \alpha D_t + (1 - \alpha) (a_{t-1} + b_{t-1})$, $b_t = \beta (a_{t-1} + b_{t-1}) + (1 - \beta) b_{t-1}$

$$t_f = -1/\lambda \ln \{R_1(t)/100\} \quad (2)$$

Where, $\lambda = \frac{1}{MTBF}$

$$L_{remain} = P_o \times (L_m - L_{used}) \quad (3)$$

Where, $P_o = \frac{1}{1 + \frac{\lambda}{\mu}}$

ii) Populate F_{t+1} , which utilizes a_t and b_t (the level estimate and trend estimate of the historical data of RVS subsystems failures and mineral concentrate demand used for the forecasting or prognosis process), in order to predict when each of the subsystems of the RVS machine will fail as well as forecast the future mineral concentrate demand of the customers through synthesis and analysis of the past failure information of the RVS subsystems and past mineral concentrates demand (D_t) using the smoothing constant of the level and trend (α and β) associated with the historical RVS subsystems failure and mineral concentrates demand data. iii) Determine, t_f , which utilizes λ and $R_1(t)$ {the failure rate, which is a function of Mean Time Between Failure (MTBF) and presumed reliability value}. This is used to determine the time at which each subsystem of the RVS machine will fail when each component reliability tends towards a threshold elastic value (i.e. reliability $< 50\%$). iv) Compute the average result of the predicted time at which the different subsystems of the RVS machine will fail, obtained from Holt’s Model and Markov Model is used to compute and ascertain when the different subsystems of the RVS machine will fail. v) Determine the remaining useful life of the RVS machine (L_{remain}), which utilizes P_o , L_m and L_{used} (steady state availability of the RVS machine; the lifespan and service life of the RVS machine). vi) Assess the steady state availability of the RVS machine is assessed using λ and μ (failure rate and repair rate) of the machine. The predicted information on when each subsystems will fail or when the mineral concentrate demand will change from the old demand to a new demand is vital in

determining the amount of RVS machine subsystems or spare parts that needs to be ordered for optimal maintenance of the machine as well as for the reconfiguration of machine required in meeting the forecasted demand.

3.3 RVS machine maintenance planning and inventory management algorithm

In order to ensure JIT maintenance of the faulty subsystems of the RVS machines and JIT reconfiguration of the machine to meet up with customers ever-changing demands as well as mark up the production loss that emanates during machine breakdown and maintenance, intelligent RVS machine inventory algorithms are required. The step by step approach to this task are: i) Utilization of a lean supply chain system, which exploits a “make to order” strategy, to produces just the amount of spare part that is needed for the maintenance and reconfiguration of the RVS machine, when the need arises. iia) Formulation of an inventory model for the RVS machine management. This model is highlighted as follows:

The following parameters are used for the model formulation.

t_1 = RVS machine subsystem(s) or spare part(s) order time required to produce Q_i

t_2 = RVS subsystem(s) or spare part(s) production time to produce Q_i

t_3 = RVS subsystem(s) or spare part(s) quality control or inspection time to produce Q_i

t_4 = RVS subsystem(s) or spare part(s) rework time (due to defects) to produce Q_i

t_5 = RVS subsystem(s) or spare part(s) transportation time to haul Q_i to the users of the RVS machine

t_6 = RVS subsystem(s) or spare part(s) loading time

t_7 = RVS subsystem(s) or spare part(s) unloading time

Using these parameters, iia) compute the total amount of quantities (Q_i) ordered by the users of RVS machine. Q_i is determined using eqn (4).

$$Q_i = \sum_{i=1}^n RVS - Sub_i \quad (4)$$

Where i denotes the number of RVS machine subsystems or spare parts that is required for the maintenance of the machine or the reconfiguration of the RVS machine in meeting the new customers demand.

iib) Calculate the lead time (L_{t-ord}) required to receive the RVS machine spare parts as well as replace worn out subsystems of the RVS machine or to reconfigure the RVS machine, provided that there is no rework during JIT manufacturing of the RVS machine spare parts using eqn (5).

$$L_{t-ord} = t_1 + t_2 + t_3 + t_4 + t_5 + t_6 + t_7 \quad (5)$$

iic) If there is re-work during the manufacturing of the spare parts needed by the users of the RVS machine, calculate the total lead time required to receive the RVS machine spare parts as well as replace worn out subsystems of the RVS machine or to reconfigure the RVS machine using eqn (6).

$$L_{t-ord} = t_1 + t_2 + 2t_3 + 2t_4 + t_5 + t_6 + t_7 \quad (6)$$

iid) Determine the total lead time (TL_t) required in replacing all the worn out subsystems of the RVS machine or reconfiguring the RVS machine using eqn (7).

$$TL_t = L_{t-ord} + L_{t-replace} \quad (7)$$

Where, $L_{t-replace}$ is the total time required to replace the worn out subsystems on the RVS machine.

Please note that $L_{t-order}$ can also be referred to as the repair time in a situation where the faulty subsystem need to be repaired.

iie) On one hand, considering this lead time as a measure to determine the specific capacity at which the machine must be configured to recovering this production loss, compute the reconfiguration time (R_t) required to recover the aforementioned production loss using eqn (8).

$$R_t = \frac{\text{Total production loss (TPL) during the period of RVS machine maintenance}}{\text{Productivity Gain (PG) of the RVS machine}} \quad (8)$$

Where, $TPL = TL_t \times \text{Productivity Capacity (PC) of the RVS machine at its first configuration}(PC_{RVS \text{ at } 1st \text{ Config.}})$, and $PG = PC_{RVS \text{ at } nth \text{ Config.}} - PC_{RVS \text{ at } 1st \text{ Config.}}$.

Note that 'n' is the new configuration status of the RVS machine in recovering the production loss.

iif) On the other hand, considering this lead time as a measure to determine the specific capacity at which the machine must be configured to produce the new customers mineral concentrate demand, calculate the reconfiguration time (R_t) required to meet new customers demand using eqn (9).

$$R_t = \frac{\text{Productivity gain-expectant } (P_{\text{gain-expectant}})}{\text{Productivity Gain (PG) of the RVS machine}} \quad (9)$$

Where, $P_{\text{gain-expectant}} = \text{Current Mineral Concentrates Demand (CD)} - \text{Old Mineral Concentrates Demand (OD)} + \text{Total Production Loss (TPL) that emanates during the reconfiguration of the RVS machine.}$

Based on the aforementioned inventory equations or algorithms, the time required to reconfigure the machine in order to mark-up for the production loss that might emanate during the maintenance of the RVS machine as well as meeting new customers mineral concentrates demand can be determined in ensuring that the production targets and customer demands of the company utilizing RVS machines are met.

4. Conclusion

RVS is an innovative beneficiation machine with high potential to be used in different surface and underground mines in the global world, for screening of different sizes of mineral particles and volumes which is required by the customer through the geometric transformation of the side plates of the machine. To this effect, intelligent decision support algorithms or models such as intelligent RVS machine diagnosing algorithms, prognosing algorithms and maintenance planning and inventory algorithms; that would ensure its optimal maintenance and management system when used in the mining industries were extensively discussed in the paper. The RVS machine diagnosing algorithms are required by the maintenance manager of the RVS machine to determine when the different subsystems of the RVS machine will fail while the RVS machine prognosing algorithms will be used to predict and forecast when the different subsystems of the RVS machine will fail as well as predict the future customers mineral concentrates demand as a yardstick to predict when and what time to reconfigure the RVS machine. Also, the maintenance planning and inventory algorithm is used to determine the amount of spare parts required for maintaining and reconfiguring the RVS machine as well as ascertaining the reconfiguration time of the RVS machine in marking the production loss during RVS machine maintenance and meeting new customers mineral concentrates demand. Successful application of this methodology would for the first time formulate a knowledge base advisor that could be used to optimally maintain RVS machine as well as effectively manage the beneficiation of varieties of mineral concentrates (in terms of sizes and volumes) needed or demanded by the customers.

References

- Adeyeri MK, Kareem B, (2012) Maintenance Dynamics, Tools for Machines Functionality in a Competitive Environment. *Industrial Engineering Letters* 2(7), pp. 12 – 19.
- Al-Najjar B, Alsyouf I (2003) Selecting the most efficient maintenance approach using fuzzy multiple criteria decision making. *International Journal of Production Economics* 84(1), pp. 85–100.
- Ashayeri J (2007) Development of computer –aided maintenance resource planning (CAMRP): A case of multiple CNC machining centers. *Robotics and Computer-Integrated Manufacturing* 23, pp. 614-623.

- Chan FTS, Lau HCW, Ip RWL, Chan HK, Kong S (2005) Implementation of total productive maintenance: A case study. *International Journal of Production Economics* **95**, pp. 71-94
- Cholette ME, Djurdjanovic D (2014) Degradation modeling and monitoring of machines using operation-specific hidden Markov models, *IEEE Transactions* **46**(10), pp. 1107-1123, DOI: 10.1080/0740817X.2014.905734
- Cui L, Li H (2006) Opportunistic maintenance for multi-component shock model. *Mathematical Methods of Operations Research* **63**.
- Dhillon BS, Liu Y (2006) Human Error in Maintenance: A Review. *Journal of Quality in Maintenance Engineering* **12**(1), pp. 21-36.
- Dhillon BS (2006) Maintainability, Maintenance, and Reliability for Engineers. *Taylor and Francis*.
- Holgado M, Macchi M (2014) Exploring the role of e-maintenance for value creation in service provision. *20th ICE Conference – IEEE TMC Europe Conference*, Bergamo, Italy.
- Lee J (2001) A framework for next-generation E-maintenance system. *2nd International Symposium on Environmentally Conscious Design and Inverse Manufacturing*, Tokyo, Japan.
- Lepojević V, Anđelković-Pešić M (2011) Forecasting Electricity Consumption by using Holt-Winters and Seasonal Regression Models. *Economics and Organization* **8**(4), pp. 421 – 431.
- Maalal S, Addou M (2011) A new approach of designing Multi-Agent Systems. *International Journal of Advanced Computer Science and Applications* **2**(11), pp. 148 – 157.
- Mori M, Fujishima M (2013) Remote monitoring and maintenance system for CNC machine tools. *8th CIRP Conference on Intelligent Computation in Manufacturing Engineering, Procedia CIRP* **12**, pp. 7-12.
- Mourtzis D, Vlachou E, Milas N, Xanthopoulos N (2016) A cloud-based approach for maintenance of machine tools and equipment based on shop-floor monitoring. *48th CIRP Conference on Manufacturing Systems* **41**, pp. 655 – 660.
- Muller A, Crespo Marquez A, Iung B (2008) On the concept of e-Maintenance: review and current research. *Reliability Engineering and System Safety* **93**(8), pp. 1165– 1187.
- Rajput HS, Jayaswal P (2012) A Total Productive Maintenance (TPM) Approach to Improve Overall Equipment Efficiency. *International Journal of Modern Engineering Research* **2**(6), pp.4383–4386.
- Ramatsetse BI (2014) Development of a reconfigurable vibrating screen. Master's Dissertation, Tshwane University of Technology, Pretoria, South Africa.
- Sheu DD, Kuo JY (2006) A model for preventive maintenance operations and forecasting. *Journal of Intelligent Manufacturing* **17**, pp. 441–451, Springer, Science Business Media.
- Shrotri AP, Khandagale SB (2012) Design of a Predictive Maintenance Program. *International Journal of Engineering and Advanced Technology*, ISSN: 2249 – 8958 **1**(4), pp. 242-246.
- Telgen D, Moergestel L, Puik E, Meyer J (2012) Agile Manufacturing Possibilities with Agent Technology. *Flexible Automation and Intelligent Manufacturing*.
- Zhang W, Halang W, Diedrich C (2003) An agent-based platform for service integration in e-maintenance. *Proceedings of IEEE international conference on industrial technology* **1**, pp. 426–33.

Robotic System Methodology for Automated Maintenance of a Reconfigurable Vibrating Screen.

O. A. Makinde*, K. Mpofu**, B. I. Ramatsetse*** and M. K. Adeyeri****

*Department of Industrial Engineering, Tshwane University of Technology, Pretoria, South Africa
olasumbomakinde@gmail.com

**Department of Industrial Engineering, Tshwane University of Technology, Pretoria, South Africa
mpofuk@tut.ac.za

***Department of Industrial Engineering, Tshwane University of Technology, Pretoria, South Africa
RamatsetseBI@tut.ac.za

****Department of Mechanical Engineering, Federal University of Technology, Akure, Nigeria
adeyerimichaeltut@gmail.com

Abstract. Reconfigurable Vibrating Screen (RVS) machine is an innovative beneficiation solution designed at Tshwane University of Technology (TUT) for beneficiating mineral particles of different sizes and volumes demanded by the customers through geometric transformation of its structure for both surface and underground mines. In order to ensure optimal and safe maintenance of different subsystems of the RVS machine, when utilized in underground mines and reduce the exposure of RVS machine maintenance managers to dangerous and hazardous underground working environment, there is a need to develop an intelligent robotic solution, that could be used to automatically maintain and manage the RVS machine at different configurations depending on the mine users' needs. In view of this, this paper proposes a novel methodology that could be used to automatically maintain and manage the RVS machine. This methodology unfolds the different elements such as maintenance tasks identification system, therblig and morphological systems, conceptual robotic end effector system, kinematic and dynamic transformation systems as well as control scheme and system, required to develop the robotic system for automatically maintaining the RVS machine, when used in an underground mine environment.

1. Introduction

The demand and requirements of mineral concentrate users has become unpredictable, dynamic and stochastic over the years while mining and beneficiation machine breakdown has been inevitable contingencies that has resulted to imbalance and unstable mineral concentrates productivity as well as hindering most of the mining companies in meeting their production targets (Makinde, 2014). To address this aforementioned problem, Reconfigurable Vibrating Screen (RVS) machine which utilizes Reconfigurable Manufacturing System (RMS) principles developed by Koren (1999), is an innovative beneficiation solution designed to screen different mineral particle sizes and volumes (in meeting the dynamic mineral concentrates demand of the customers and marking-up the production loss that could emanate from the machine breakdown) through geometric transformation of the machine structure. This innovative solution envisaged to solve the aforementioned problem is foreseen to be utilized in different surface and hazardous underground mines. To this effect, different maintenance strategies such as corrective maintenance; "fix it when it is broken" technique (William, 1999), preventive maintenance; "fix it before it fail" technique (Badia *et al.*, 2002 and Liao *et al.*, 2010), opportunistic maintenance; "fix opportuned components that is subjected to failure" technique (Samhoury *et al.*, 2009 and Tambe and Kulkarni, 2013), predictive maintenance; "use of diagnosing and prognosing algorithms" (Faiz and Edirisinghe, 2009); dynamic maintenance; "customer-demand induced maintenance planning technique (Adeyeri and Kareem, 2012) and e-maintenance management system; "use of Information

Communication Technology (ICT) and real-time monitoring sensors strategy” (Muller *et al.*, 2008) e.t.c. needs to be optimally blended in order to optimally maintain and manage RVS machines, when used in surface and underground mining environment. However, to complement this maintenance management of the RVS machine, in order to ensure its safe and danger-free maintenance; when used in dangerous and hazardous underground mining environment, there is a need to develop an intelligent robotic solution that is capable of automatically maintaining and reconfiguring the RVS machine when the need arise in an underground mining environment. In view of this, this aim paper is aimed at developing a methodological framework which could be used to develop the intelligent robotic solution required for automated maintenance of the RVS machine. This framework unfolds the RVS machine maintenance tasks identification technique, therblig and morphological systems, conceptual robotic end effector system, robotic kinematic and dynamic transformation systems and control scheme and system required by the intelligent robotic system solution for automated maintenance of the RVS machine solution. In light of this, the first section of this paper presents the state of art in the application of robotic systems for the maintenance of industrial machines. The second section of this paper discusses the Reconfigurable Vibrating Screen (RVS) machine while the last section of this paper discusses the methodological framework for developing an intelligent robotic solution customized for automatically maintaining RVS machine, when used in underground mines.

2. State of the art in the application of robotic systems in industrial machine maintenance.

According to Andrew *et al.*, 2008; Zhang and Chu, 2010 and Farnsworth *et al.*, 2015, maintenance is a strategic decision and activities carried out on a machine in order to ensure its optimal functionality at its designed level of reliability and safety. From Farnsworth *et al.*, 2015 perspective, industrial machine maintenance activity is traditionally viewed as a dull, dangerous or dirty operation as well as a non-productive function during the manufacturing of a product. To this effect, strategic maintenance policies required for maintaining industrial machines used in dangerous and hazardous working environment or whose maintenance tasks are dull and ergonomically challenging using human being has called for the development of a robotic-driven maintenance system. To this effect, Schempf *et al.*, 2003 designed an autonomous robot that could be used for inspecting oil and gas pipelines for crack leakage and detection. Bo *et al.*, 2010 also designed an autonomous robot, which utilizes Magnetic Flux Leakage (MFL) technique to carry out a real-time inspection of fatigue and cracks on aircraft delivery tubes in an aerospace industry. Luk *et al.*, 2005 designed an autonomous robot that could be used in welding ship hulls as well as inspecting power plants. Jiang and Wistort (2005) designed an electromechanical and sensing system that could be used in controlling autonomous robot used for inspecting electric power cables. Based on the aforementioned robotic maintenance solution successes, one could affirm that tapping or exploitation of the speed, power, availability and improved accuracy of robot systems in machine maintenance have significantly promoted safe, accurate and optimal maintenance of machines used in different manufacturing industries. To this effect, it therefore seems logical to explore the possibilities of automating the maintenance of different subsystems of the RVS machine when used in underground mining environment in order to exploit the same benefits enjoyed by the afore-discussed robotic maintenance solutions.

3. Reconfigurable Vibrating Screen

Reconfigurable Vibrating Screen is an innovative beneficiation solution that can be used for screening different mineral particle sizes and volumes that are dynamically demanded by the customers in both surface and underground mining environment (Ramatsitse, 2014). This machine performs its screening

operation through the relative functional actions of its different subsystems such as screen panels, torsion bar, torsion bar bracket, side plates, side liner plates, back plates, back liner plates, screen deck frame, vibrating motors, bolts and bearings e.t.c. while its functionality seizes when any of the aforementioned subsystem of the RVS machine fails. In view of this, there is a need to develop an intelligent robotic solution that could be used to automatically maintain and manage these different faulty subsystems of the RVS machine; when the need arise in an underground mining environment.

4. Holistic methodological framework for automated maintenance of a RVS machine using an intelligent robotic solution.

The methodological framework for the automated maintenance of the RVS machine is made of different modules such as the maintenance task identification system, therblig and morphological systems, conceptual robotic end effector technique, robotic-driven kinematic and dynamic transformation systems and robotic-driven control scheme and system as depicted in Figure 1.

4.1 Maintenance task identification system: The intelligent robotic-driven maintenance system controller is equipped with an interface which communicates to it via a wireless communication protocol; on when the different parts of the RVS machine needs to be maintained using an intelligent diagnosing algorithms of the RVS machine maintenance management system. The diagnosing algorithms are made up of ruled-based algorithms formulated by RVS machine experts using intelligent diagnosing systems, which utilizes intelligent sensors and the aforementioned rules for real-time machine data acquisition and maintenance decision making.

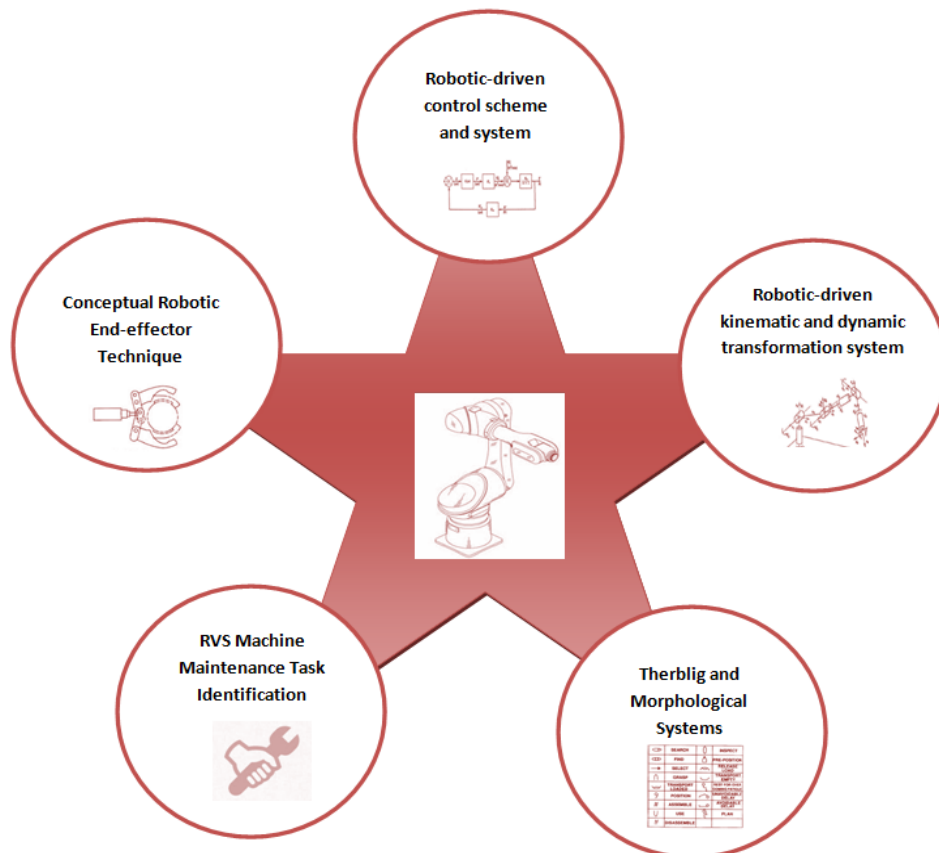


Figure 1: Robotic System Methodology for automated maintenance of a Reconfigurable Vibrating Screen

4.2 Therblig and morphological systems: According to Jia *et al.*, 2013, therblig is a system used to assess and evaluate the different types of motions required in performing a task while the morphology is a scientific approach used to study and ascertain how different objects are been formed or how different tasks are carried out in a work environment (Onwuama *et al.*, 2012). Therblig algorithms are vital for automated maintenance of RVS machine because the different set of motions that a human being will use in maintaining the RVS machine can be determined as a yardstick to ascertain the trajectory motions that the robot needs to exhibit in automatically maintaining the different subsystems of the RVS machine while the morphological algorithms are important in order to ascertain the work functions that must be carried out by human being in maintaining the RVS machine, which thus serve as a yardstick in ascertaining the functional parameters required by the intelligent robotic system for automated maintenance of the RVS machine. Furthermore, the therblig and morphological methodologies can also be used to train the intelligent robotic-driven maintenance solution via a robot teach pendant; which studies, assesses and analyzes these motion records and activities obtained from these methodologies. For example, the therblig and morphological methodologies required to replace a worn out screen panel on the RVS machine can be obtained by studying how human being remove and replace worn-out screen panels on faulty conventional vibrating screens. To this effect, a detailed step by step activities carried out by human being in removing and replacing worn out subsystems of the RVS machine is highlighted as follows:

Step 1 (S1): Removal of faulty screen panels on the deck frame of the RVS machine

S1a) Loose or unpin the screen panel pins that are holding the screen panel to deck frame using a screen panel pin removal tool.

S1b) Remove (using your hands), the loosened pins and place it in a warehouse bin A, where new and healthy subsystems of the RVS machine are kept before usage. If the loosened pin has worn-out, place it in warehouse bin B, where faulty subsystems of the RVS machine are kept before it is transferred to the recycling companies for recycling purposes.

S1c) Remove by hand, the loosened worn-out screen panels and transport it to the warehouse bin B, where it is kept before it is transferred to the recycling companies for recycling purposes.

Step 2 (S2): Place the new screen panels obtained from the warehouse bin A (where new and healthy subsystems of the RVS machine are kept before usage) on the screen deck frame.

Step 3 (S3): Hammer the loosened screen panel pins that was obtained in the warehouse panel A into the screen panel pins holes in order to firmly hold the screen panel pins to the screen panel deck frame using a hammering tool.

Therefore, the therblig algorithms required by the intelligent robotic solution to remove and replace worn-out subsystems of the RVS machine is depicted in Table 1.

Table 1: Therblig Algorithm for automated removal and replacement of worn-out screen panels on the RVS Machine

Maintenance Task	Therbligs	Therblig Symbols
S1a	S1a: Loose or unpin	#
S1b	S1b: Remove (i.e. hold) + Grasp + Transport Loaded + Place (i.e. pre-position and position)	
S1c	S1c: Remove (i.e. hold) + Grasp + Transport Loaded + Place (pre-position and position)	
S2	S2: Search + find + hold + grasp + transport loaded + place (pre-position and position)	

S3

S3: Hammer (i.e. Assemble)

#

Also, the morphological algorithm required by the intelligent robotic solution to remove and replace worn-out subsystems of the RVS machine is depicted in Table 2.

Table 2: Morphological Algorithm for automated removal and replacement of worn-out screen panels on the RVS Machine

Maintenance Task	Function	Function Type	Robotic-driven Solution Approach	Elaborated Robotic solutions
S1a	Loose or unpin	Action-driven	Force Analysis	Force
S1b	Remove (i.e. hold) + Grasp + Transport Loaded + Place (i.e. pre-position and position)	Action-driven	Electrostatic Force and Transferred Kinematic Energy.	Force + Workdone + Power + Object Recognition
S1c	Remove (i.e. hold) + Grasp + Transport Loaded + Place (pre-position and position)	Action-driven	Electrostatic Force and Transferred Kinematic Energy.	Force + Workdone + Power + Object Recognition
S2	Search + find + hold + grasp + transport loaded + place (pre-position and position)	Cognitive and Action-Driven	Vision and Object Recognition System, Electrostatic Force and Transferred Kinematic Energy.	Vision system + Object Recognition + Force + Workdone + Power
S3	Hammer (i.e. Assemble)	Action-driven	Force Analysis	Force

4.3 Conceptual robotic end effector system: in order to achieve the automated maintenance of the RVS machine, an end effector, equipped with the ability to grip and transport different subsystems of the RVS machine; loose and bolts the M-20 bolts holding the different subsystems of the RVS machine, as well as un-pin and hammer the screen panel pins holding the screen panels of the RVS machine to the screen deck frame is required. To this effect, a multifunctional end effector, equipped with a worm and gear module for gripping and grasping actions, a crow bar module for unpinning and loosening of the screen panel pins and a customized reduced-scale electronic-controlled cylinder for hammering the screen panels is proposed as a suitable end effector for the intelligent robotic solution that will be used in automatically maintaining the RVS machine.

4.4 Robot-driven kinematic and dynamic transformation system: the kinematic and dynamic transformation system formulates the joint angles, force and torque required by the robotic manipulator of the intelligent robotic solution in automatically maintaining the different subsystems of the RVS machine. The trajectory or kinematic movement of the robot customized in automatically maintaining the different subsystems of the RVS machine is ascertained by determining the relative joint angles or torques that must be supplied by the servo motor controllers of the robot system, through a detailed inverse kinematic analysis of the spatial position in x, y and z coordinates of the different subsystems to be maintained on the RVS machine, obtained from intelligent laser scanners using inverse kinematic algorithms formulated by Gou *et al.* (2010). The relative forces and torques required in maintaining the different subsystems of the RVS machine will be ascertained by using the forward and inverse dynamic algorithms formulated by Saha *et al.* (2014). Furthermore, the four-wheel drive autonomous robot customized for automatically maintaining the RVS machine moves from its initial position to final position (where it needs to maintain the different of subsystems of the RVS machine) using the linear and angular velocities of the robot wheels that is estimated from the autonomous robot wheel control model and equations formulated by Tlale and Villiers (2008).

4.5 Robot-driven control scheme and system: this is an important phase of the intelligent robotic system which controls how a robot automatically maintains the different subsystems of the RVS machine. A detailed control scheme for the robot system customized for automatically maintaining RVS machine in an underground mining environment is depicted in Figure 2. In this scheme, the laser scanning sensors and vision sensors maps the underground mining environment to determine the location (i.e. the position and orientation) of different subsystems of the RVS machine as well as obstacles (static and dynamic) situated within the path or route where the machine is located. This information will then be sent to the autonomous robotic wheel speed and encoder sensors, robotic manipulator joint torque sensors and end effector force and torque sensors to ascertain the linear and angular speed/ acceleration required by the robot wheels in maneuvering to desired position in the underground mining environment as well as ascertain the amount of force and torque that is required by the intelligent autonomous robotic system manipulators to reach these aforementioned positions on the RVS machine during its maintenance using its kinematic and dynamic algorithms described in previous section of this paper. The established forces and torque obtained from the aforementioned sensors is then sent to microprocessor of the robot control for further processing and coding into unified control rules needed by the servo-controllers of the robot system. The servo-controllers send the necessary command to the robot actuators in order to set the robot wheel, robot arm and robot end effector in the necessary motions required in carrying out different maintenance tasks on the RVS machine. For precise control of the robot system customized for automated maintenance of the RVS machine, errors obtained during the control of the robot wheels, robot arm and robot end effectors are normalized using servo-regulators and amplifiers of the robot controller. Furthermore, since the automated maintenance tasks achievement is an iterative activity, the aforementioned control sequences runs for the number of maintenance activity motion iterations required for the maintenance tasks to be completed on the RVS machine.

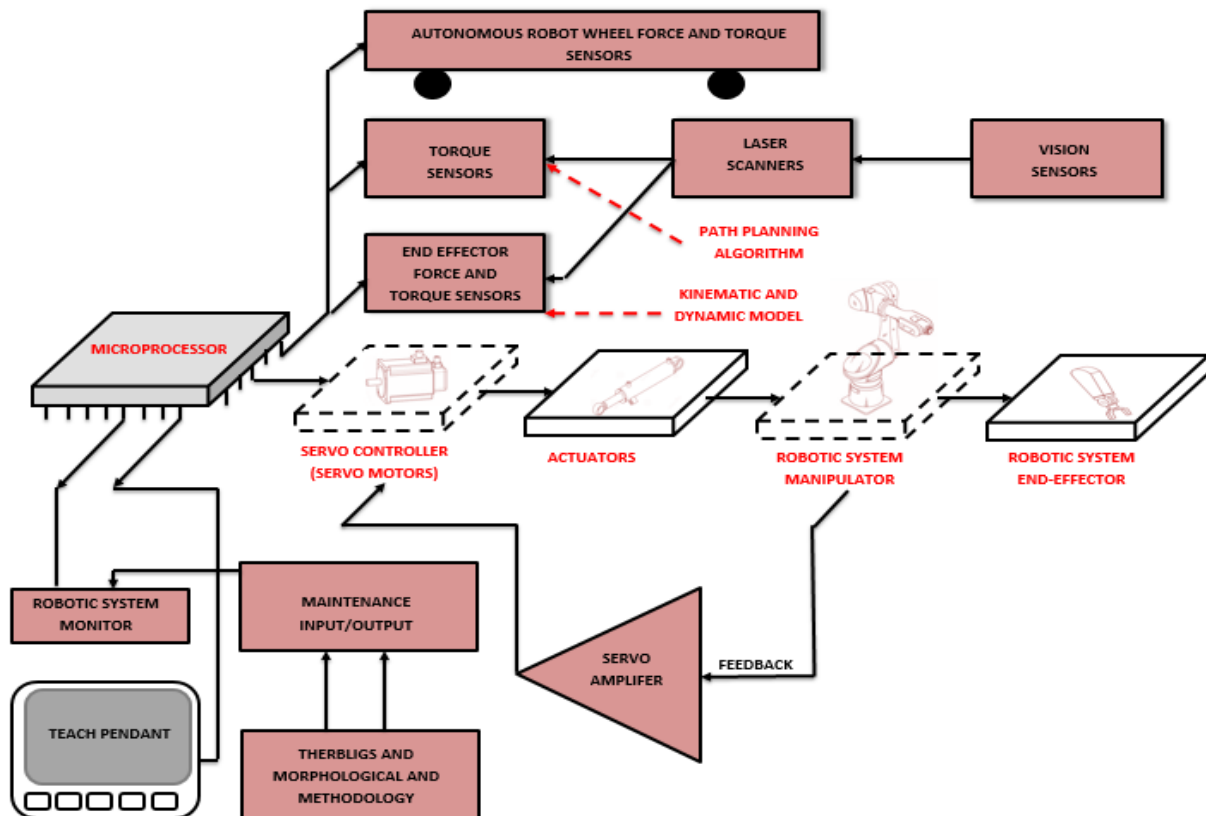


Figure 2: Control scheme and system for automated maintenance of a Reconfigurable Vibrating Screen

5. Conclusion

Reconfigurable Vibrating Screen (RVS) machine is the next line of beneficiation technology designed in Tshwane University of Technology, South Africa; for both surface and underground mines, to screen different mineral particle sizes and volumes that are stochastically demanded by the customers. In order to ensure optimal and safe maintenance of the RVS machine when utilized in dangerous and hazardous underground mines, robotic-driven maintenance solution was proposed as a potential solution for safely maintaining the RVS machine, when used in an underground mine. In light of this, a methodological framework which holistically discusses the robot maintenance task identification system, robotic-driven maintenance therblig and morphological systems, conceptual robot end effector system, robotic-driven maintenance kinematic and dynamic transformation system as well as robot-driven maintenance control scheme and system required for optimal functionality and mobility of the robotic-driven maintenance solution in a hazardous and dangerous underground mining environment have been extensively discussed. The robot maintenance task identification system, therblig and morphological systems and conceptual robotic end effector system reveals the techniques to be used by the robot in identifying when and how to replace the different subsystems of the RVS machine using a robotic solution as well as the relevant conceptual tool that must be used by the robot in automatically maintaining the RVS machine. The kinematic and dynamic transformation system and the robot control scheme and system reveals how the robot maneuver within the underground mine environment in reaching the different subsystems of the RVS machine to be maintained and also reveals the control sequences followed by the controller of the robotic solution in automatically maintaining the different subsystems of the RVS machine. This holistic methodological framework provides the relevant information and parameters needed by the robotic maintenance solution designers in developing an intelligent robotic solution required for automatically maintaining the RVS machine.

References

- Adeyeri MK, Kareem B (2012) Maintenance Dynamics, Tools for Machines Functionality in a Competitive Environment. *Industrial Engineering Letters* 2(7), pp. 12 – 19.
- Andrew T, Byard P, Henry P (2008) Self maintenance works for repair firm. *Engineering Technology*, 3(7), pp. 69–72.
- Badia FG, Berrade MD, Campos CA (2002) Optimal inspection and preventive maintenance of units with revealed and unrevealed failures. *Reliability Engineering and System Safety*, 78(2), pp. 157–163. DOI:10.1016/S0951-8320(02)00154-0.
- Bo L, Feilu L, Zhongqing J, Jiali, L (2010) Eddy current array instrument and probe for crack detection of aircraft tubes. *International conference on intelligent computation technology and automation*.
- Faiz RB, Edirisinghe EA (2009) Decision Making for Predictive Maintenance in Asset Information Management. *Interdisciplinary Journal of Information, Knowledge, and Management* 4, pp. 1 – 14.
- Farnsworth M, Bell C, Khan S, Tomiyama T (2015) Autonomous Maintenance for Through-Life Engineering. Book Chapter, Springer International Publishing Switzerland, L. Redding and R. Roy (eds.), *Through-life Engineering Services, Decision Engineering*, DOI 10.1007/978-3-319-12111-6_23, pp. 1 – 26.
- Gou Z, Sun Y, Yu H (2010) Inverse Kinematics Equation of 6-DOF Robot Based on Geometry projection and simulation. *International Conference on Computer, Mechatronics, Control and Electronic Engineering (CMCE)*, IEEE 978-1-4244-7956-6, pp. 125 – 128.

- Jia S, Tang R, Lv J (2013) Therblig-based energy demand modeling methodology of machining process to support intelligent manufacturing. *Journal of Intelligent Manufacturing*, DOI 10.1007/s10845-012-0723-9, pp. 1 – 19.
- Jiang B, Wistort M (2005) Autonomous Robotic Monitoring of Underground Cable Systems. ICAR 05, *Proceedings of 12th International Conference on Advanced Robotics*, Seattle, WA.
- Liao W, Pan E, Xi L (2010) Preventive maintenance scheduling for repairable system with deterioration. *Journal of Intelligent Manufacturing* **21**, pp. 875–884.
- Luk BL, Cooke DS, Galt S, Collie AA, Chen S (2005) Intelligent legged climbing service robot for remote maintenance applications in hazardous environments. *Robotics and Autonomous Systems* **53**, pp. 142–152.
- Makinde OA (2014) Functionality Assessment of a Reconfigurable Vibrating Screen. *Master's Degree Thesis*, Tshwane University of Technology, South Africa, pp. 1-266.
- Muller A, Crespo Marquez A, Iung B (2008) On the concept of e-maintenance: review and current research, *Reliability Engineering and System Safety* **93**, pp. 1165-1187.
- Onwuama KT, Salami SO, Ali M, Nzalak JO (2012) Effect of different methods of bone preparation on the skeleton of the African Giant Pouched Rat (*Cricetomys gambianus*). *International Journal of Morphology* **30**(2), pp. 425-427.
- Ramatsetse BI (2014) Development of a reconfigurable vibrating screen. *Master's Dissertation*, Tshwane University of Technology, Pretoria, South Africa.
- Saha SK (2014) 3D Modeled-Based Robotics Learning Software User Manual. <http://www.roboanalyzer.com>, pp. 1-21.
- Samhoury MS, Al-Ghandoor A, Fouad RH, Alhaj Ali SM (2009) An Intelligent Opportunistic Maintenance (OM) System: A Genetic Algorithm Approach. *Jordan Journal of Mechanical and Industrial Engineering* **3**(4), pp. 246 – 251.
- Schempf H, Mutschler E, Goltsberg V, Crowley W (2003) Grislee: gas main repair and inspection system for live entry environments. *International Journal of Robotic Research* **22**(7–8), pp. 603–616.
- Tambe PP, Kulkarni MS (2013) An Opportunistic Maintenance Decision of A Multi-Component System Considering the Effect of Failures on Quality. *Proceedings of the World Congress on Engineering* **1**, WCE 2013, London, U.K, pp. 1-6.
- Tlale N, Villiers M (2008) Kinematics and Dynamics Modelling of a Mecanum Wheeled Mobile Platform. *15th International conference on Mechatronics and Machine Vision in Practice*, ISBN: 978-0-473-13532-4, pp. 657-662.
- Zhang Z, Chu X (2010) A new approach for conceptual design of product and maintenance. *International Journal of Computer Integrated Manufacturing* **23**, pp. 603–618.

Modelling imperfect repair in a degradation system. A stochastic comparison

S. Mercier*, I.T. Castro**

* *Laboratoire de Mathématiques et de leurs Applications, Batiment IPRA, Avenue de l'Université, F-64013, Université de Pau et des Pays de l'Adour, Pau (France): sophie.mercier@univ-pau.fr*

** *Escuela Politécnica, University of Extremadura, Cáceres (Spain): inmatorres@unex.es*

Abstract We consider two different imperfect repair models applied to a degrading system. In the first model, the imperfect repair reduces the degradation accumulated in the system since the last maintenance action. In the second model, the repair reduces the additional age accumulated in the system from the last maintenance action. The aim of this presentation is to compare these two imperfect repair models for a non-continuously monitored system with periodic maintenance times. For that, we assume that the degradation of the system follows a non-homogeneous gamma process.

1. Introduction

A classical stochastic model to describe a non-decreasing continuous accumulated random degradation is the gamma process. The gamma process is a stochastic process with independent, non-negative increments having a gamma distribution with identical scale parameter. If $X(t)$ represents the degradation of the system at time t and is modelled using a gamma process with parameters $\Lambda(t)$ and β , then for any t and Δ , the increment $X(t + \Delta) - X(t)$ is a gamma distributed random variable with parameters $\Lambda(t + \Delta) - \Lambda(t)$ and β .

This gamma deterioration process was successfully applied to model condition-based preventive maintenance specially when the deterioration level of the system is only collected through inspections (Noortwijk 2009, Kahle *et al.* 2016). In these cases, preventive maintenance tasks are key elements to slow the degradation process of the system and reduce the corrective maintenance actions.

In many cases, the maintenance tasks restored the system to an “as good as new condition”. However, in a general setting, the adjustments performed in the system lower the system deterioration but they do not bring the system back to its brand new state. They hence correspond to imperfect maintenance tasks. Although imperfect maintenance in the context of degradation-based maintenance has not received much attention until now, some imperfect maintenance models have been developed (Zhang *et al.* 2015, Mercier and Castro 2013, Castro and Mercier 2015).

In this presentation, we compare two types of imperfect maintenance for a degrading system modelled as a non homogeneous gamma process. The first type of maintenance reduces the system degradation accumulated from the last maintenance action. The second one reduces the age of the system and, consequently, its degradation. The comparison of the two imperfect repair models is performed considering some stochastic orders such as the increasing convex ordering. The results of this analysis are applied to compare the accumulated expected reward of the system in an interval of time considering the two types of imperfect maintenance. Numerical examples illustrate the analytical results.

2. The two models of imperfect repairs

Let $\{X_t, t \geq 0\}$ be a non homogeneous gamma process with parameters $\Lambda(t)$ and β , that is, the probability density of X_t is given by

$$f_{\Lambda(t),\beta}(x) = \frac{\beta^{\Lambda(t)}}{\Gamma(\Lambda(t))} x^{\Lambda(t)-1} e^{-\beta x}, \quad x \geq 0,$$

where $\Lambda(t)$ is a non-decreasing and continuous function with $\Lambda(0) = 0$.

Let $X^{(i)}$ be independent and identical copies of the process $\{X_t, t \geq 0\}$. We assume that an imperfect repair is performed each T units of time. The development of the process depends on the type of repair performed in the maintenance times. We consider two types of imperfect repair: a reduction of the system degradation and a reduction of the system age.

2.1. First model: Arithmetic Reduction of Deterioration of order 1 (ARD1)

We assume that a brand new system starts working at time $t = 0$. An imperfect repair is performed each T units of time. This imperfect repair reduces in a $\rho\%$ the degradation accumulated by the system since the last maintenance action (or since $t = 0$) where $0 < \rho < 1$. The maintenance action is considered instantaneous.

Let $\{Y_t, t \geq 0\}$ be the process that describes the degradation of the maintained system. We have

$$Y_t = X_t^{(1)}, \quad t < T; \quad Y_T = (1 - \rho)X_T^{(1)}$$

For $T \leq t < 2T$, we get

$$Y_t = Y_T + (X_t^{(2)} - X_T^{(2)}), \quad T \leq t < 2T; \quad Y_{2T} = Y_T + (1 - \rho)(X_{2T}^{(2)} - X_T^{(2)})$$

In a general setting, for $nT \leq t \leq (n+1)T$ and $n = 1, 2, \dots$

$$Y_t = Y_{nT} + (X_t^{(n+1)} - X_{nT}^{(n+1)}), \quad nT \leq t < (n+1)T; \quad Y_{(n+1)T} = Y_{nT} + (1 - \rho)(X_{(n+1)T}^{(n+1)} - X_{nT}^{(n+1)}).$$

It is easy to show that the expectation and the variance of Y_t are given by

$$\mathbb{E}(Y_t) = \frac{\Lambda(t) - \rho\Lambda(n_t T)}{b}, \quad \text{Var}(Y_t) = \frac{\Lambda(t) - \rho(2 - \rho)\Lambda(n_t T)}{b^2},$$

where n_t corresponds to the integer part of t/T .

2.2. Second model: Arithmetic Reduction of Age of order 1 (ARA1)

We assume that an imperfect repair is performed each T units of time. This imperfect repair reduces in a $\rho\%$ the age accumulated by the system since the last maintenance action (or since $t = 0$) where $0 < \rho < 1$. The maintenance action is considered instantaneous.

Let $\{Z_t, t \geq 0\}$ be the process that describes the degradation of the maintained system. We have

$$Z_t = X_t^{(1)}, \quad t < T \quad Z_T = X_{(1-\rho)T}^{(1)}.$$

For $T \leq t < 2T$, we get

$$Z_t = Z_T + (X_{t-\rho T}^{(2)} - X_{(1-\rho)T}^{(2)}), \quad T \leq t < 2T; \quad Z_{2T} = Z_T + (X_{2(1-\rho)T}^{(2)} - X_{(1-\rho)T}^{(2)}).$$

In a general setting, for $nT \leq t \leq (n+1)T$ and $n = 1, 2, \dots$

$$Z_t = Z_{nT} + (X_{t-\rho nT}^{(n+1)} - X_{(1-\rho)nT}^{(n+1)}), \quad nT \leq t < (n+1)T; \quad Z_{(n+1)T} = \sum_{i=1}^{n+1} (X_{(1-\rho)iT}^{(i)} - X_{(1-\rho)(i-1)T}^{(i)}).$$

It is easy to show that the expectation and the variance of the process Z_t are given by

$$\mathbb{E}(Z_t) = \frac{\Lambda(t - \rho n_t T)}{b}, \quad \text{Var}(Z_t) = \frac{\Lambda(t - \rho n_t T)}{b^2},$$

where n_t corresponds to the integer part of t/T .

In the ARA1 model and, in case of a non homogeneous gamma process, the imperfect repair not only reduces the deterioration level of the system at maintenance times but, in the common case of an increasing rate of deterioration, it also reduces its age and consequently its rate of deterioration. In case of a non homogeneous gamma process, both models hence deeply differ and they will not model imperfect repairs with similar maintenance actions. (In one case, only the deterioration level is reduced whereas in the other case, the system is additively rejuvenated).

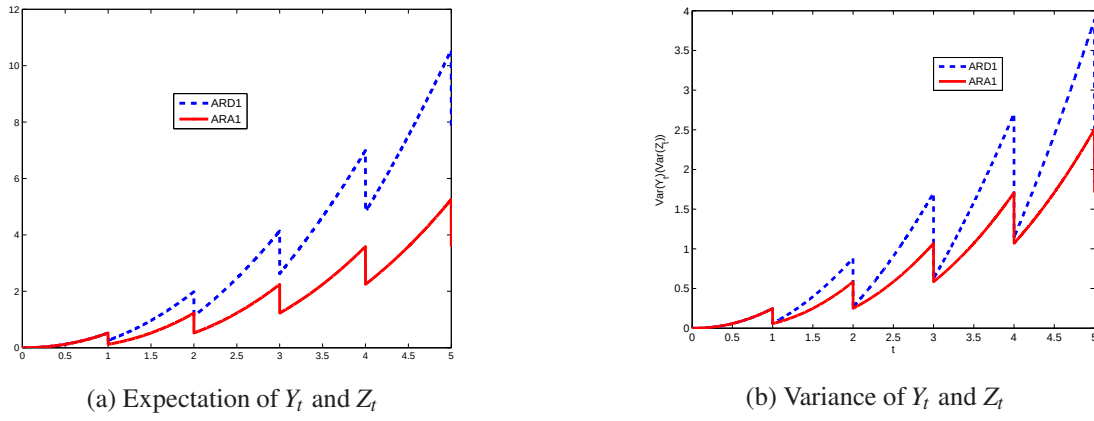


Figure2. Expectation and variance in the convex case

Figure 1 shows a realization of the two processes Y_t and Z_t . This realization has been performed for $\Lambda(t) = \alpha t^\beta$ where $\alpha = 1.1$ and $\beta = 2$ with imperfect repairs performed each $T = 1$ units of time with a repair efficiency of $\rho = 0.5$.

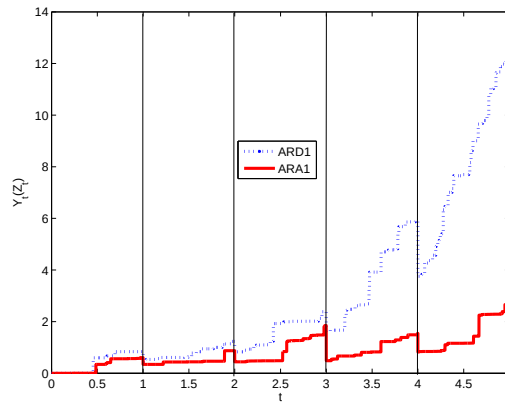


Figure1. A realization of Y_t and Z_t

3. Some comparisons between the two models

A usual problem in probability and statistics is to compare the position and variability of two distributions. The simplest way of doing it is by the comparison of the mean and standard deviations. Now, we compare the moments of the processes Y_t and Z_t taking into account the concavity and convexity of the function $\Lambda(t)$. So, we get

Theorem 1. *If the function $\Lambda(t)$ is concave, then*

$$\mathbb{E}(Y_t) \leq \mathbb{E}(Z_t), \quad \text{var}(Y_t) \leq \text{var}(Z_t), \quad \forall t \geq 0.$$

If the function $\Lambda(t)$ is convex and fulfills

$$\Lambda(at) \leq a^2 \Lambda(t), \quad \forall a \in [0, 1], \quad \forall t \geq 0, \quad (1)$$

then,

$$\mathbb{E}(Z_t) \leq \mathbb{E}(Y_t), \quad \text{var}(Z_t) \leq \text{var}(Y_t) \quad \forall t \geq 0.$$

Figure 2 shows the expectation and the variance of the processes Y_t and Z_t calculated using the function $\Lambda(t) = \alpha t^\beta$ where $\alpha = 1.1$, $\beta = 2.1$ with an efficiency repair of $\rho = 0.5$ and assuming that the system is repaired each $T = 1$ time units. We can check that, for this dataset, inequality (1) is fulfilled. Hence, from Theorem 1 the expectation and the variance for the ARD1 case are greater than the expectation and the variance for the ARA1 case.

Figure 3 shows the expectation and the variance of the processes Y_t and Z_t calculated using the function

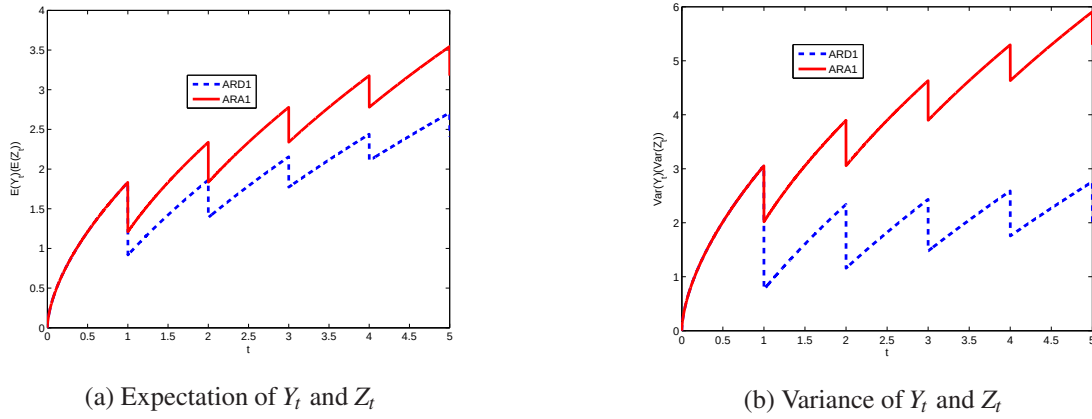


Figure3. Expectation and variance in the concave case

$\Lambda(t) = \alpha t^\beta$ where $\alpha = 1.1$, $\beta = 0.6$ with an efficiency repair of $\rho = 0.5$ and assuming that the system is repaired each $T = 1$ time units. From Theorem 1 the expectation and the variance for the ARA1 case are greater than the expectation and the variance for the ARD1 case.

In the literature, several orders which compare the variability of two distributions functions have been defined. One of these is the increasing convex ordering (see Shaked and Shantikumar 2007 for definitions and more details).

DEFINITION 2. Given two random variables X e Y with distribution functions F_X and F_Y , X is smaller than Y in the increasing convex (concave) order, and it is denoted by $X \leq_{icx} Y$ ($X \leq_{icv} Y$), if

$$\int_t^\infty \bar{F}_X(x)dx \leq \int_t^\infty \bar{F}_Y(x)dx, \quad \left(\int_0^t F_X(x)dx \geq \int_0^t F_Y(x)dx \right) \forall t.$$

An important result about increasing convex order is that $X \leq_{icx} Y$ if and only if $E(\phi(X)) \leq E(\phi(Y))$ holds for any non-decreasing convex real function ϕ for which the expectations are defined. In the same way, $X \leq_{icv} Y$ if and only if $E(\phi(X)) \leq E(\phi(Y))$ holds for all non-decreasing concave real function ϕ for which the expectations are defined.

We get the following result

Theorem 3. If $\Lambda(t)$ is convex, then $Z_t \leq_{icv} Y_t$ for all $t \geq 0$. If $\Lambda(t)$ is concave, then $Y_t \leq_{icx} Z_t$.

In the homogeneous case, that is $\Lambda(t) = \alpha t$, the rate of deterioration remains unchanged in time so that both models boil down to some deterioration reduction at maintenance times. In the homogeneous case, it is easy to check that

$$\mathbb{E}(Y_t) = \mathbb{E}(Z_t).$$

Hence, assuming homogeneous case, we can compare the two process considering other stochastic ordering.

DEFINITION 4. Given two random variables X e Y , X is said to be smaller than Y in the convex (concave) order (and it is denoted by $X \leq_{cx} Y$) if

$$\mathbb{E}(\phi(X)) \leq \mathbb{E}(\phi(Y)),$$

holds for any real convex function ϕ for which the expectations are defined.

Theorem 5. If the gamma process is homogeneous, that is, $\Lambda(t) = \alpha t$, we have that $Z_t \leq_{cx} Y_t$ for all $t \geq 0$.

4. Expected accumulated reward of the system in an interval of time

We assume that a new system is put in operation at time $t = 0$. This system is subject to a continuous degradation modeled as a non homogeneous gamma process with parameters $\Lambda(t)$ and β . This system is inspected at periodic times $T, 2T, 3T, \dots$ and an imperfect repair (ARA1 or ARD1) is performed in the inspection times. We also consider that the system provides an unitary reward function (per unit time) which is higher when the deterioration level of the system is lower. The reward function is assumed to be decreasing (and continuous).

Also, we consider that, from a degradation level c , the rate of reward decreases more quickly. Based on classical functions used in the insurance literature (Rolski *et. al* 1999), we consider the following (unitary) reward function:

$$g(x) = (b_1 - e^{\alpha_1 x}) \mathbf{1}_{\{0 \leq x \leq c\}} + (b_2 - e^{\alpha_2 x}) \mathbf{1}_{\{c \leq x\}}, \quad x \geq 0, \quad (2)$$

with $b_1, b_2, \alpha_1, \alpha_2, c > 0$ and $b_2 = b_1 + e^{\alpha_2 c} - e^{\alpha_1 c}$, $\alpha_1 \leq \alpha_2$.

It is easy to check that the function g given by (4) is continuous, non-increasing and concave. Figure 4 shows the behaviour of function g . As we can check, for values greater than $x = 4$, the utility function decreases more quickly. Furthermore, for values greater than $x = 10$, function g is negative.

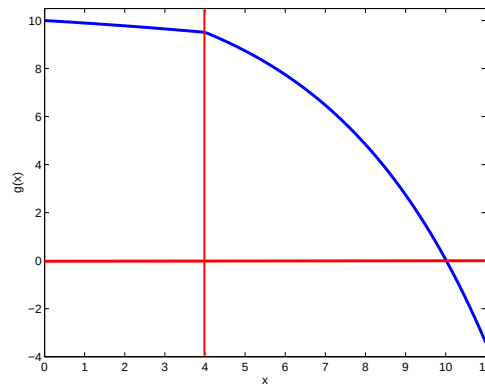


Figure4. Function reward

Under this framework and in absence of failures, the expected accumulated reward in the time interval $[0, t_0]$ for both types of maintenance is given by

$$r_Y(t_0) = \int_0^{t_0} \mathbb{E}[g(Y_s)] ds, \quad r_Z(t_0) = \int_0^{t_0} \mathbb{E}[g(Z_s)] ds,$$

for a ARD1 model (AR1 model) respectively.

Theorem 6. *If $\Lambda(t)$ is concave, then*

$$r_Y(t_0) \geq r_Z(t_0).$$

Figure 5 shows the expected accumulated reward function for the two models. This example has been calculated as follows. We consider a system subject to a degradation modelled as a non-homogeneous gamma process with parameters $\Lambda(t) = \alpha t^\beta$ where $\alpha = 2$ and $\beta = 0.8$. We assume the system works according to the reward function given by (4) with $\alpha_1 = 0.1$, $\alpha_2 = 0.25$, $c = 4$ and $b_1 = 11$. The repair efficiency is given by $\rho = 0.5$. Figure 5 shows the expected accumulated reward at time $t_0 = 10$ versus T . The data have been obtained using Monte Carlo simulation for 9 equally spaced values from 0.5 to 10 with 10000 simulations in each point to compute the expectation ($\mathbb{E}[g(Y_s)]$ and $\mathbb{E}[g(Z_s)]$) and, after, the Simpson's rule is applied to evaluate the integral. Based on Figure 5 and according to Theorem 3, the expected accumulated reward for ARD1 model is greater than the expected accumulated reward for ARA1 model.

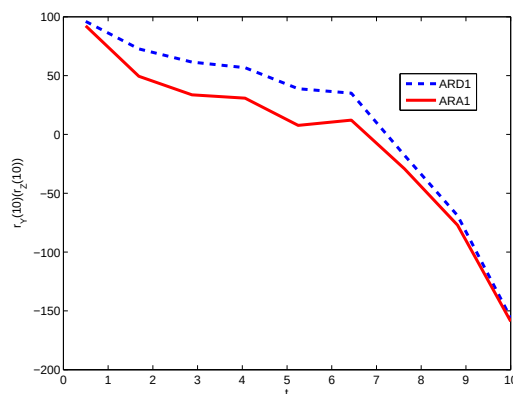


Figure5. Expected accumulated reward at time $t_0 = 10$ versus T

5. Conclusions

In this presentation, we describe two stochastic processes obtained by applying two models of imperfect repair to a degradation system (ARD1 and ARA1). Some properties about the stochastic orderings of the two processes are showed. Under some assumptions and in absence of failures, we prove that the expected accumulated reward using the ARD1 model is greater than the expected accumulated reward using the ARA1 model. This property is illustrated with a numerical example. A maintenance strategy including preventive maintenance actions and corrective maintenance actions will be analyzed in future works.

Acknowledgments

For I.T. Castro, the research was supported by *Ministerio de Economía y Hacienda*, Spain (Project MTM2015-63978-P) and *Gobierno de Extremadura*, Spain (Project GR15106) partially funded by the European Regional Development Fund (ERDF).

References

- van Noortwijk JM (2009) A survey of the application of gamma processes in maintenance. *Reliability Engineering and System Safety*, 94(1), 2-21.
- Kahle W, Mercier S, and Paroissin C (2016) *Degradation Processes in Reliability*. John Wiley & Sons, *under press*.
- Zhang M, Gaudoin O, Xie M (2015) Degradation-based maintenance decision using stochastic filtering for systems under imperfect maintenance *European Journal of Operational Research*, 245(2), 531-541.
- Mercier S, Castro IT (2013) On the modelling of imperfect repairs for a continuously monitored gamma wear process through age reduction. *Journal of Applied Probability*, 50(4), 1057-1076.
- Castro IT, Mercier S (2015) Performance measures for a deteriorating system subject to imperfect maintenance and delayed repairs. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, to appear.
- Shaked M, Shantikumar J.G. (2007) *Stochastic Orders*. Springer.
- Rolski T., Schmidli H., Schmidt V., Teugels J. (1999) *Stochastic Processes for Insurance and Finance*. Wiley.

Shock and damage model parameter estimation and forecasting with an unobserved number of shocks

N. Montgomery*, D. Banjevic**

*Centre for Maintenance Optimization and Reliability Engineering, University of Toronto, neilm@mie.utoronto.ca

**Centre for Maintenance Optimization and Reliability Engineering, University of Toronto, banjev@mie.utoronto.ca

Abstract We investigate the problem of parameter estimation and forecasting for a shock and damage model in which the number of shocks cannot be observed. Shock and damage models are widely studied by reliability theorists but there has been little published work on the statistical analysis of datasets related to such models. It is usually assumed that the number of shocks and the amounts of damage are observed to some extent. A company asked us to analyse a dataset in which only the total amount of damage (possibly zero) was observed. We proposed a Poisson process model for the number of shock events and estimated the rate parameter using the method of maximum likelihood, obtaining numerical estimates for the rate parameter and a standard error for the estimate. We then estimated the average amount of damage per shock event using a method of moments estimator. We used these parameter estimates to forecast the probability of one or more total failures among the entire population of all such fittings. The company was able to use our results and forecasts to make key asset management decisions relating to this class of fittings.

1. Introduction

Items of any kind can, and do, fail as a result of damage due to one or more sudden external forces that cause the item to exceed a threshold beyond which it is capable of operating. Often the occurrence of the external force or the amount of damage are unpredictable to the extent that they are modelled using random processes. So shock and damage models, and similar models under different names, have received plenty of attention in the academic literature in pure and applied probability, reliability, insurance, and others.

The basic idea is usually that the shocks occur as a homogeneous Poisson process $N(t)$ with rate λ . The cumulative shock times T_i therefore have Erlang distributions with parameters λ and i . The amount of damage X_i suffered at shock i is considered to be constant or modelled by some non-negative random variable. The total damage suffered up to time t is $Z(t) = \sum_{i=1}^{N(t)} X_i$, a doubly stochastic process. The total damage eventually causes the item to fail (which censors the underlying process).

This model is accessible to the novice reliability theorist. Extensions have been developed in all possible probabilistic directions, considering any variety of shock models, damage models, various dependencies among all the processes, and the effects all such variations might have on the first passage time properties that might characterize the actual moment of failure. From all such models a wide variety of maintenance policies might be developed. The literature in this area is vast and we do not attempt to summarize it here. In the field of reliability we point the interested reader in the direction of Nakagawa (2007) and its extensive list of references as a place to start.

There is relatively little discussion of the problem of parameter estimation for all the proposed models. Kahle and Wendt (2010) list three possible data scenarios. They do so in the context of a model in which a frailty variable Y is included permitting the underlying shock process to vary by individual item, which we can ignore, reducing the number of data scenarios to two. In the first scenario, the shock and degradation history (T_i, X_i) is observed, and the underlying parameters for each process can be estimated separately. This scenario seems to be implicitly assumed in most practical work in shock and damage models, which might explain the lack of attention given to parameter estimation. In the second scenario only the failures are observed, a situation also treated in Murthy and Iskander (1991).

We encountered a third scenario when a company asked us to help them with a situation they faced in which they were able to determine, through destructive testing, only the total damage done to an asset class (of “fittings”) but not the actual number of shocks (unless the number is 0). The assets in operation were buried underground and the shock mechanism was believed to be the freeze-thaw cycle of the soil which was not directly observed. The company had no insights that might inform a possible distribution for the amount of

damage suffered during each shock. They knew how much total damage an item could accumulate before total failure, which as an added complication was a function of each item's initial state upon installation possibly many decades ago. This initial state is equivalent to an initial amount of "damage". No total failure had ever occurred. The company was seeking forecasts for the probability of one or more total failures occurring in the population of fittings. They had received forecasts from the testing firm hired to do the physical analysis of the items and wanted our review of their calculations.

In this paper we implement a very simple shock and damage model but in what appears to be a novel situation in which no shocks are observed, no failures occurred, and only the total amount of damage (possibly 0) per item is available in a sample. We use the model to make useful forecasts of failure probability for the population.

2. Data description, model specification, and parameter estimation

2.1. Data

The dataset consisted of data collected from n destructive tests undertaken on a sample of fittings that had been exhumed for the express purpose of assessing the health of the population of interest. The two variables of interest at the parameter estimation stage of the analysis are the raw total amounts of damage suffered $d^{\text{raw}} = \{d_1^{\text{raw}}, \dots, d_n^{\text{raw}}\}$ and the ages $t = \{t_1, \dots, t_n\}$. Another measurement was taken to estimate the item's initial state, but this variable is only needed for forecasting later on.

It is unknown how many shock events, if any, each item had endured. We only know if the item had 0 events, or 1 or more events. Define the variable d based on d^{raw} by letting $d_i = 0$ when $d_i^{\text{raw}} = 0$ and $d_i = 1$ otherwise.

2.2. Shock and damage model

To model the shock process we use a homogeneous Poisson process model with parameter λ for the (mostly) unobserved shock history and estimate λ using maximum likelihood as follows. The number of shocks endured by item i at time t is denoted by $N_i(t)$. So the probabilities of item i having endured either 0 or 1 or more shocks during its lifetime t_i are given by:

$$P(N_i(t_i) = 0) = e^{-\lambda t_i}, \quad (1)$$

and

$$P(N_i(t_i) > 0) = 1 - e^{-\lambda t_i}. \quad (2)$$

The likelihood for λ is therefore:

$$L(\lambda) = \prod_{i=1}^n \left(e^{-\lambda t_i} \right)^{1-d_i} \left(1 - e^{-\lambda t_i} \right)^{d_i}, \quad (3)$$

with log-likelihood given by:

$$l(\lambda) = -\lambda \sum_{i=1}^n t_i (1 - d_i) + \sum_{i=1}^n d_i \log \left(1 - e^{-\lambda t_i} \right). \quad (4)$$

There is no closed-form solution to find the maximum of this function, but it is straightforward to accomplish using widely available routines. We use the *bbmle* package in R (Bolker et al. 2014) which produces estimates and standard errors. As a first approximation to $\hat{\lambda}$, required by the numerical routine, we use $\sum_{i=1}^n d_i / \sum_{i=1}^n t_i$, in the spirit of a total-time-on-test style of estimator.

The company had no insight into the possible distribution of the damage suffered at a shock event. We choose to treat the damage as a constant, acknowledging that this tends to result in conservative projections of failure probabilities. The constant amount of damage a nevertheless needs to be estimated, which we do indirectly by considering the product λa , which is the average amount of damage per unit time. We use a method of moments approach to the estimation of λa by simply dividing the total amounts of damage by the total item ages to obtain the following estimator:

$$\widehat{\lambda a} = \frac{\sum_{i=1}^n d_i^{\text{raw}}}{\sum_{i=1}^n t_i}. \quad (5)$$

To estimate a itself we compute the maximum likelihood estimator $\hat{\lambda}$ for λ and solve for a in equation (5) to

obtain:

$$\hat{a} = \frac{\widehat{\lambda a}}{\hat{\lambda}}. \quad (6)$$

The company's destructive testing subcontractor had itself attempted to estimate λ and a . They used an interesting but ultimately doomed ad-hoc method, beginning with a rather than λ . To estimate a they used the average of the non-zero damages d_i^{raw} . This is a sensible first approximation that might be accurate if shocks are so rare that most items would suffer only 0 or 1 events, but given the variation in the d_i^{raw} in the data this was unlikely to be the case, certainly if the damage is to be treated as a constant as we do in this paper. We believe the subcontractor's estimate for a was therefore probably too large. Their estimate for λ involved imputing the number of shock events for each item from their average damage estimate and using these imputed values as "data" to plug into the standard MLE for λ from a Poisson process. Since their a was probably too large their λ was probably too small.

To compare our estimates with the subcontractor's estimates we performed a variety of simulations which showed ours to be as accurate as the maximum likelihood theory gives. The subcontractor's estimates tended to overestimate a and underestimate λ unless λ was small enough so that very few cases of 2 or more shocks occurred, as we suspected.

3. Forecasting failure probabilities

An item will fail once it accumulates sufficient total damage. In general this threshold might be fixed or random. In this case study the threshold is fixed on a per-item basis but will vary depending on an initial condition which depends on how the item was installed. The n items exhumed for the sample each had this initial condition measured, and so the amount of damage each could have suffered can be calculated. Since we are treating the per-shock damage a as constant, the probability of item i failing by time t is simply:

$$P_i(t) = 1 - \sum_{j=0}^{k_i-1} \frac{e^{-\lambda t} (\lambda t)^j}{j!}, \quad (7)$$

where k_i is the number of shocks required to cause item i to fail, which is a function of a and item i 's initial condition. The probability of having one or more failures from the sampled items before time t , had they been left in service, is given by:

$$P^{(n)}(t) = 1 - \prod_{i=1}^n (1 - P_i(t)). \quad (8)$$

The company was interested in forecasting the probability of one or more failures in the entire population of N items. This probability depends in part on λ , a , and the distributions of ages of items in the population. The ages are known and λ and a are estimated from the sample. It also depends on the distribution of initial conditions which determine the number of shocks a particular item can suffer before failure, which is not known. If we assume the initial values within the sample are representative of the population we can estimate the probability of one or more failures in the population by $1 - (1 - P^{(n)}(t))^{N/n}$. This effectively "scales up" the sample size to the population size using the sample's average failure probability for every item in the population.

4. Exploring the data and numerical results

The age and damage amounts from the original data have been rescaled to obscure the true values, with even the rescaled values omitted from plots, but the overall properties of the sample remain very similar. The sample size was $n = 212$, out of which $\sum_{i=1}^n (1 - d_i) = 136$ items had suffered no damage. Figure 1 shows the distribution of ages in years of items in the sample, which is bimodal due to the nature of the way the company had expanded and acquired smaller companies over several decades. Figure 2 shows the distribution of the amount of damage suffered by those 76 items that had suffered non-zero damage.

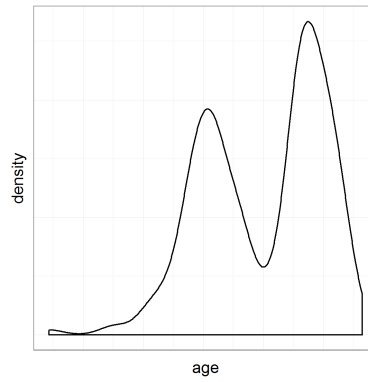


Figure 1. Density plot of ages in years of sampled items.

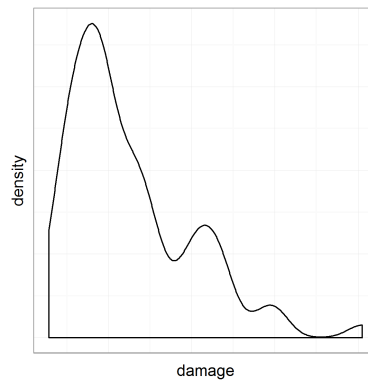


Figure 2. Density plot of total damage suffered by sampled items that had suffered any damage at all.

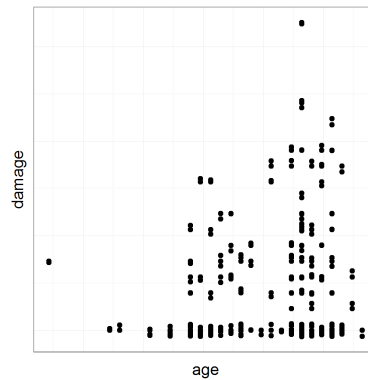


Figure 3. Scatterplot of age versus amount of damage.

Figure 3 shows the relationship between age and amount of damage suffered in items from the sample, with some vertical jitter added to points to deal with some overplotting due to the physical limits on the accuracy of per-item damage estimation in the destructive testing procedure.

4.1. Parameter estimation

The rate parameter maximum likelihood estimate is $\hat{\lambda} = 0.0090068$ with standard error 0.0010431. The estimate of the damage per unit time is $\hat{\lambda}a = 0.0358$, giving an estimate for the amount of damage per shock of $\hat{a} = 3.98$.

These values can be compared to the subcontractor's ad hoc estimates of 5.0 for the damage per shock and 0.00764 for the shock rate. The product of these two subcontractor estimates, which estimates the damage per unit time, is 0.0382. These over-estimations of a and λa compared with our more accurate calculations results in substantial over-estimation of the probabilities of one or more failures in the entire population.

Table 1. Probability of one or more failures in population within five year increments, with parameter estimates given.

Years	Low	Middle	High
$\hat{\lambda}$	0.011093	0.009007	0.006921
$\hat{a}$	3.23170	3.98023	5.18004
5	0.000000	0.000001	0.000028
10	0.000001	0.000022	0.000433
15	0.000012	0.000162	0.002144
20	0.000067	0.000661	0.006627
25	0.000244	0.001956	0.015789
30	0.000701	0.004715	0.031841
35	0.001698	0.009864	0.057097
40	0.003633	0.018579	0.093697
45	0.007067	0.032262	0.143248
50	0.012744	0.052467	0.206445
55	0.021601	0.080779	0.282706
60	0.034760	0.118643	0.369948
65	0.053485	0.167138	0.464584
70	0.079116	0.226719	0.561829
75	0.112951	0.296970	0.656310
80	0.156085	0.376416	0.742877
85	0.209206	0.462473	0.817447
90	0.272382	0.551576	0.877624
95	0.344846	0.639535	0.922945
100	0.424871	0.722069	0.954666

4.2. Failure probability forecasts

The company was ultimately interested in long-term forecasting of the probabilities of one or more items in the population suffering a total failure. If the probability is too high for the company's tolerance for risk, the likely outcome would be to consider an expensive project to replace the current population with new fittings more resistant to stress and damage. We use a population size of $N = 100000$, which is within an order of magnitude of the true value.

We use the parameter estimates from 4.1 in equations (7) and (8) to make the probability forecasts. We also use $\hat{\lambda} \pm 2 \cdot \text{std.err.}(\hat{\lambda})$ to propose simple lower and upper values on plausible forecasts (without going so far as to calling these limits "confidence intervals"). Table 1 shows probabilities of one or more failures in the population. The estimate of the damage a is very important for these projections. The "High" probability estimates actually correspond to the lower bound on the λ , which results in a higher value for $\hat{a}$ and a higher forecasted probability.

5. Discussion and future work

This is a paper about dealing with partial information. We developed maximum likelihood and method of moment estimators for the parameters of a simple shock and damage model but for the novel situation in which no shock history was available and no failures had been observed—only the total amount of damage on a sample of items was available. Had a small number of failures occurred our approach can still be used since the occurrence of a failure would indicate one more shocks had been suffered. The majority of items had no damage and it is interesting to note that having items with no damage is required for our method to be useful.

We used these estimates to produce long range probability forecasts for one or more failures in the population of assets. Our approach is capable of forecasting probabilities for any number of failures, which could be of interest for other applications in which failures are more tolerable and expected. The company who asked us to analyse the problem was able to use these forecasts to inform an important and possibly expensive asset management decision.

It would be interesting extend our approach to improve upon estimation of a , to relax the constant damage assumption, and to consider the relationship between age and damage.

References

- Bolker B, R Development Core Team (2014) *bbmle: Tools for general maximum likelihood estimation*. R package version 1.0.17. <https://CRAN.R-project.org/package=bbmle>
- Kahle W, Wendt H (2010) Parametric Shock Models In *Advances in Degradation Modeling* (M.S. Nikulin et al., Eds.), Birkhäuser Boston, pp.77–104.
- Murthy DNP, Iskander, BP (1991) A new shock damage model: Part I Model formulation and analysis. *Reliability Engineering and System Safety*, 31, 191-208.
- Nakagawa T (2007) *Shock and damage models in reliability theory*. Springer.

A Model Reduction for Probabilistic-Reliability (RPR)

Static analysis using a set of representative functions

L. Ncib**, A. Manai*, H. Hassis*, N. Gmati**

Université de Tunis El Manar, École Nationale d'Ingénieurs de Tunis,

*LR-03-ES05 Laboratoire de Génie Civil-ENIT

**LR-99-ES20 LAMSIN - ENIT

BP 37, Le Belvédère 1002, Tunis, Tunisia

Email: hedi.hassis@enit.rnu.tn

Abstract In mechanics, as in many scientific domains, information and data can be deterministic (certainty known or with error - safety coefficient), probabilistic (random with known probability distribution) or possibilistic (random known with an uncertainty factor on the reliability of the information. Parameters change in a known range, but the probability is unidentified). The probabilistic mechanical approaches aim to give an estimation of the solution when the information on the parameters of analyzed problem is uncertain. For some mechanical problems with probabilistic parameters, a constraint to be satisfied is added. In engineering design, the reliability analysis is an example of constraint. Here, and for static probabilistic-reliability analysis, an approach reducing the size of numerical computing is introduced. The solution is projected on a special set of representative functions; the number of functions is at least the number of probabilistic parameters. Only one bigger problem is solved and its dependence with probabilistic parameters is analytical. Then any optimization or analysis is obtained without any new calculation. The present approach is validated on static discrete and continuous beam systems having probabilistic parameters. Results are compared with the Monte Carlo simulations.

1. Introduction: Probabilistic analysis for the reliability

To improve the security level of a mechanical system, the reliability analysis is a permanent concern of designer engineers, mainly when the parameters are random or uncertain. While analyzing the sensitivity of a system with respect to random variables [O. Dessombz et al. (2001), R. M. Moore (1966, 1979), J. Rohn (1996), D.M. Frangopol (1985), M. Lemaire (1997), C.G. Bucher and U. Bourgund (1990)], the objective of a reliability analysis is to:

- Compare the reliability needs (under stress) and the resistance (availability of materials),
- Calculate the probability of failure in which the need is greater than or equal to the resistance.

This mechanical-reliability approach deals necessarily with two complementary models: mechanical modeling and reliability modeling; [P. Bjerager, and S. Krenk, (1989), S E. BALM (1996), A. Bouazzouni et al. (1997), N. Rabhi et al. (2011), U. Kirsch (1995)]. When all considerations are introduced, a large numerical problem is generally obtained where reducing the number of parameters is often recommended. Several strategies have been developed to reduce the number of needed samples. We distinguish the importance sampling [R.E.Melchers (1999)] which is based on design point estimation undertaken at the beginning, by carrying out some sampling and by only holding the point belonging to the failure domain, which is nearest to the origin of the normalized space. The density of the importance sample can then be estimated and centered at this point. The approximation methods are undertaken in the space of a standard Gaussian variable [A.M. Hasofer, 1974]. These methods are based on the evaluation of the limit state function by either approximating the reliability function by a tangent(FORM) [O. Ditlevsen, 1996], or by a paraboloid (SORM) [K. Breitung, 1998] at a failure point closest to the origin and calculating the probability of failure as a simple function of parameters. This work is a coupling between the condensation methods and reliability methods. It uses a special set of functions. Hence, the objective is to build a robust reliability model reducing both the uncertainties on the parameters of design and the structural parameter changes.

2. Static approach of the space-probabilistic separation approach

Even the approach is valid for any linear probabilistic problem, mechanical applications are chosen to valid this approach. Consider that each probabilistic variable V_i follows its law defined by its density distribution, and its cumulative distribution function $P_i(v_i)$ which can be inversed to $v_i(P_i)$.

Consider that a static problem containing n_p probabilistic parameters, V_i , defined by a linear problem:

$$\begin{aligned} K(v_1(P_1), \dots, v_{n_p}(P_{n_p})).X = F; \text{ noted } K(v_i(P_i)).X = F \\ K(P_1, \dots, P_{n_p}).X = F \end{aligned} \quad (1)$$

K is the stiffness matrix issue of the space discretization and contains n_p probabilistic parameters. Each probabilistic parameter number i , v_i , has its own probabilistic law. F is the force vector and also could depend of probabilistic parameters. Let Θ_j a known and chosen probabilistic functions composing a representative set of functions (it is not established that it is a base); j varies from 1 to at last n_p . In some cases, Θ_j can depend only of one parameter, and then the number of the used function is n_p . At others, Θ_j can depend of different parameters, and then the number of the used function is less than n_p ; n'_p will design the maximum number of functions.

By using the separation variables and by decomposition on the representative set of functions, Θ_j , well chosen, the solution X is written as the following decomposition:

$$X = \sum_{j=1}^{n'_p} \Theta_j X_j \quad (2)$$

Using the decomposition 2, equation 1 is written as:

$$K(P_i). \sum_{j=1}^{n'_p} \Theta_j X_j = F \quad (3)$$

The unknown vectors are X_j and will be determined by using the classical projection functions. It consists on multiplying the equation 3 by a function Θ_k and integrated in the probabilistic domain (from 0 to 1). This projection is expressed by equation 4.

$$\int_0^1 \int_0^1 \left[K(P_i). \sum_{j=1}^{n'_p} \Theta_j X_j \right] \Theta_k dP_1 \dots dP_{n_p} = \int_0^1 \int_0^1 [F] \Theta_k dP_1 \dots dP_{n_p}; \quad k = 1 \text{ to } n'_p \quad (4)$$

After numerical integration (at many cases, the integration is formal and analytical one), n'_p related problems are written as expressed by 5:

$$\sum_{j=1}^{n'_p} \bar{K}_{kj} X_j = \bar{F}_k; \quad k = 1 \text{ to } n'_p \quad (5)$$

The n'_p problems defined by equation 5 are written as:

$$\begin{bmatrix} - & - & - & - & - & - \\ \bar{K}_{k1} & - & - & \bar{K}_{kj} & - & \bar{K}_{kn_p} \\ - & - & - & - & - & - \\ - & - & - & - & - & - \\ - & - & - & - & - & - \\ - & - & - & - & - & - \end{bmatrix} \begin{bmatrix} X_1 \\ \cdot \\ \cdot \\ X_j \\ \cdot \\ X_{n'_p} \end{bmatrix} = \begin{bmatrix} * \\ \bar{F}_k \\ * \\ * \\ * \\ \bar{F}_{n'_p} \end{bmatrix} \quad (6)$$

$$\text{With: } \bar{K}_{kj} = \int_0^1 \int_0^1 [K(P_i) \Theta_j \cdot \Theta_k] dP_1 \dots dP_{n_p}; \quad \bar{F}_k = \int_0^1 \int_0^1 [F(P_i) \Theta_k] dP_1 \dots dP_{n_p} \quad (7)$$

The dimension of $\bar{K}_{kj}$ is the same of K. The X_j are solved and the probabilistic solution is obtained by the recombination defined by:

$$X = \sum_{j=1}^{n_p} \Theta_j X_j \quad (8)$$

Remarks: compared to a direct approach, and suppose that the probabilistic domain (between 0 and 1) is divided into only 10 intervals, for each probabilistic parameter, the reduction of the size of calculation is:

- $n_p=2$: reduction: 92% (from 100 problems with $n*n$ to 1 problem with $2n*2n$)
- $n_p=4$: reduction: 99,36 % (from 10000 problems with $n*n$ to 1 problem with $4n*4n$)
- $n_p=k$: reduction: $\left(1 - \frac{k^3}{10^k}\right) \cdot 100\%$ (from 10^k problems with $n*n$ to 1 problem with $kn*kn$)

3. Probabilistic Discrete system Applications

Notation: k is a stiffness, m is a mass, f is a force and x is a local displacement. The probabilistic parameters are defined by their mean or expectation of the distribution and the standard deviation. Their probability (cumulative distribution function) P(k) is then estimated and its inverse function k(P) is calculated. For discrete systems, solutions are computed using the maple tools.

3.1. A Two-DOF discrete system with two different probabilistic laws

Consider two degrees of freedom masse-spring system, shown in figure 1. Let k_1 be a uniform distribution law defined by its limits: $a_{k1}=8E9$ N/m and $b_{k1}=3.2E10$ N/m, and let k_2 be a Gaussian probabilistic law defined by its mean $\mu_{k2}= 1E10$ N/m, and its deviation $\sigma_{k2}= 3E9$ N/m. The forces are $f_1=f_2= 3.5E7$ N.

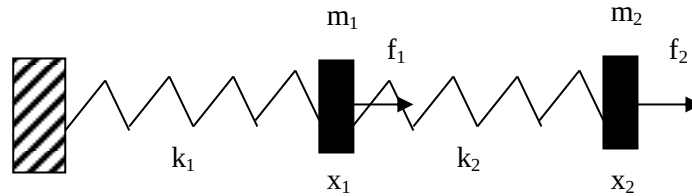


Figure 1. A two degrees of freedom (DOF) system with 2 probabilistic parameters

The static equilibrium equations are written as follows:

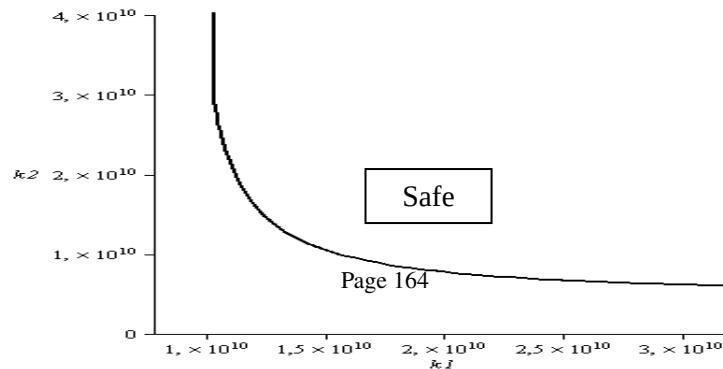
$$\begin{bmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \text{ or } K.X = F \quad (9)$$

Two functions are used here and defined by:

$$\Theta_1(P_1) = \frac{(a_{k1} + b_{k1}P_1)}{2k_1(P_1)}; \quad \Theta_2(P_2) = \frac{\mu_{k2}}{k_2(P_2)} \quad (10)$$

The relative error between the analytical solution and RPR solution is around 1E-6. As example of reliability criterion, the displacement x_2 will be less than 0.008 m. The safe and unsafe areas are represented as a function of stiffness (figure 2).

Remark: for deterministic parameters, it is necessary to choose a special function for all deterministic parameters.



Unsafe

Figure 2. safe and unsafe areas as a function of the of stiffness

3.2. A frame example

The implementation of the approach was performed on the Matlab numerical code, especially for a 2D beam finite element defined by its probabilistic parameters: the Young modulus, E , and the two section parameters a and b ; figure 3. Let a frame structure constrained at A and loaded at B by a force equal to $F=5E5$ N, as indicated in figure 3.

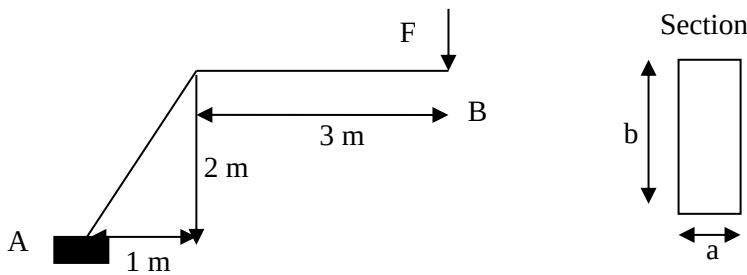


Figure 3. The analyzed structure

It is assumed that the reliability is defined by a criterion on the deflection W , at the node B, which should be smaller than a limited displacement $w_{\max} = -0.3\text{m}$. The objective function can be defined as follows:

$$G = W - w_{\max} \quad (11)$$

The objective is to calculate the probability of default, P_f , defined by:

$$P_f(G = W - w_{\max} < 0) \quad (12)$$

Three cases are analyzed, as defined by Table1.

Case 1	Case 2	Case 3
(E,a) probabilistic	(E,b) probabilistic	(E,a,b) probabilistic
$\mu(E)= 2E11 \text{ N/m}^2$ $\sigma(E)= 2E10 \text{ N/m}^2$ $\mu(a)= 0.035 \text{ m}$; $\sigma(a)= 0.0035 \text{ m}$ $b=0.35$	$\mu(E)= 2E11 \text{ N/m}^2$ $\sigma(E)= 2E10 \text{ N/m}^2$; $\mu(b)= 0.35 \text{ m}$; $\sigma(b)= 0.035 \text{ m}$; $a=0.035$	$\mu(E)= 2E11 \text{ N/m}^2$ $\sigma(E)= 2E10 \text{ N/m}^2$; $\mu(a)= 0.035 \text{ m}$ $\sigma(a)= 0.0035 \text{ m}$; $\mu(b)= 0.35 \text{ m}$ $\sigma(b)= 0.035 \text{ m}$
Representative set functions		
$\Theta_1(P_E, P_a) = \frac{\mu_E \mu_a}{E(P_E) a(P_a)}$	$\Theta_1(P_E, P_b) = \frac{\mu_E \mu_b}{E(P_E) b(P_b)}$ $\Theta_2(P_E, P_b) = \frac{\mu_E (\mu_b)^3}{E(P_E) b^3(P_b)}$	$\Theta_1(P_E, P_a, P_b) = \frac{\mu_E \mu_a \mu_b}{E(P_E) a(P_a) b(P_b)}$ $\Theta_2(P_E, P_a, P_b) = \frac{\mu_E \mu_a (\mu_b)^3}{E(P_E) a(P_a) b^3(P_b)}$

Table1. Definition of different analyzed cases

The results obtained by this approach are compared with those obtained by Monte Carlo (MC) simulations. Monte Carlo methods (or Monte Carlo experiments) are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results.

As a selected result, Figures .4.a and 4.b show the limit separating the area where the criterion is verified and the area where it is not verified, as a function of cumulative distribution functions of the probabilistic parameters of case 1 and case 2. The explicit result obtained by our approach is compared with the Monte-Carlo simulation.

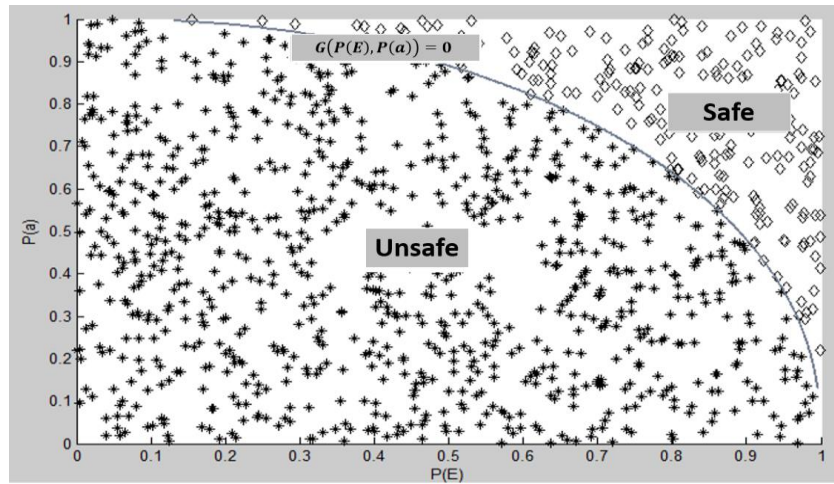


Figure 4.a Case 1: Safe and unsafe areas

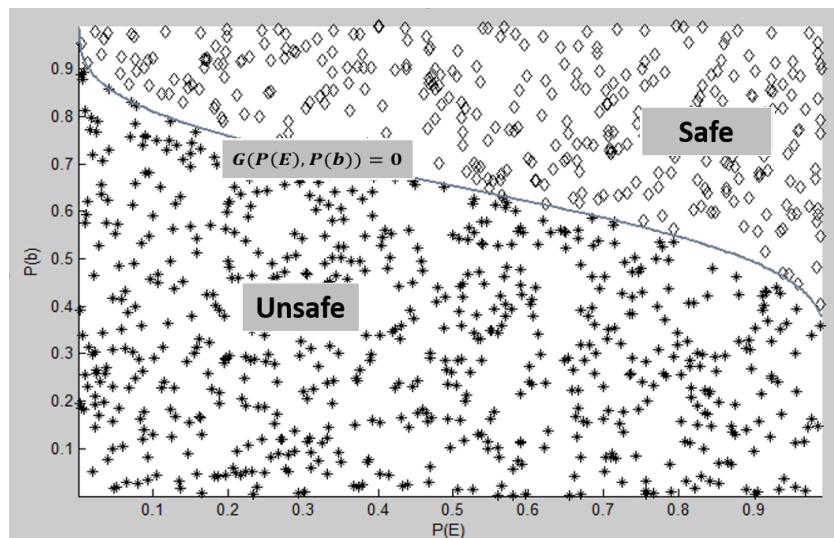


Figure 4.b Case 2: Safe and unsafe areas

Remark: From examples, and it is not demonstrated, the best set representative functions are obtained by the independent function obtained from the inverse of the elementary matrix. For any another finite element, it is suggested to inverse the elementary finite element matrix (the inversed part) and the independent functions will be so the representative set functions.

4. Conclusions and perspectives

For a reliability analysis with probabilistic parameters, classical approaches lead to a problem with a large size. Reducing its size without any alteration of the solution is recommended for a good design.

Here presented a general theory for static approach for reliability probabilistic analysis. For static problem, the solution is projected on well chosen representative set of functions and an approximated solution is obtained. The equivalent problem reduces the size problem.

The Reduction Probabilistic Reliability (RPR) approach, here described, is applied here for static probabilistic discrete and frame examples. Discrete and frame static systems are analyzed and compared with exact solutions or issued from Monte-Carlo pulling analysis.

As perspectives, these approach can be extended to possibilistic evolution parameters and to dynamic analysis in frequencies domain.

Acknowledgements:

This work is supported by the European MOBIDOC/PASRI program, ESI Tunisia, ESI France, Astrium.

References

- S E. BALM, Parametric families of reduced finite element models. Theory and applications, *Mechanical Systems and Signal Processing*, vol. 10(4), p.381-394, 1996.
- P. Bjerager, and S. Krenk, (1989). "Parametric Sensitivity in First Order Reliability Theory." J. Eng. Mech., 115(7), 1577–1582
- A. Bouazzouni, G. Lallement, S. Cogan, Selecting a Ritz basis for reanalysis of frequency response functions of modified structures, *Journal of Sound and Vibration*, vol. 199, N . 2, p. 309-322, 1997.
- C.G. Bucher and U. Bourgund. A fast and efficient response surface approach for structural reliability problems. *Structural Safety*, 7:57 66, 1990.
- K. Breitung, F. Casciati, L. Faravelli, Reliability based stability analysis for actively controlled structures, *Engineering Structure* 20 (3) (1998) 211–215.
- O. Dessombz, F. Thouverez, J.-P. La n , L. Jézéquel, Analysis of mechanical systems using interval computations applied to finite element methods, *Journal of Sound and Vibration* 239(5) 949-968, 2001.
- O. Ditlevsen, H.O. Madsen, *Structural Reliability Methods*, Wiley, Chichester, UK, 1996.
- D.M. Frangopol, Structural Optimization Using Reliability Concept. *Journal of Engineering Structures*, ASCE, v. 111, p. 2288-2301, 1985.
- A.M. Hasofer, N.C. Lind, Exact and invariant second-moment code format (an exact and invariant first order reliability format), *Journal of Engineering Mechanics* 100 (1974) 111–121.
- U. Kirsch, S. Liu, Exact structural reanalysis by a first order reduced basis approach, *Structural optimisation* 10, 153-158, Springer-Verlag, 1995.
- M. Lemaire. Reliability and mechanical design. *Reliability Engineering and System Safety*, 55 :163-170, 1997.
- R.E.Melchers, *Structural Reliability Analysis and Prediction*, second ed., Wiley, 1999.
- R. M. Moore. Interval Analysis. Prentice-Hall, 1966.
- R. M. Moore. Method and applications of Interval Analysis. *Studies in Applied Mathematics (SIAM)*, 1979.
- N. Rabhi, M. Guedri, H Hassis and N. Bouhaddi. Structure dynamic reliability: A hybrid approach and robust meta-models, *Mechanical Systems and Signal Processing*, Volume 25, Issue 7, October 2011, Pages 2313-2323
- J. Rohn. On overestimations produced by the interval gaussian algorithm. *Technical Report 690, Institute of Computer Science, Academy of Sciences of the Czech Republic*, oct 1996.

Adjusting Condition Monitoring Quality for Condition-Based Maintenance Optimization of Deteriorating Systems

T.P.K Nguyen⁽¹⁾, K.T. Huynh⁽¹⁾, P. Do⁽²⁾, C. Bérenguer⁽³⁾, A. Grall⁽¹⁾

⁽¹⁾ICD, ROSAS, LM2S, Univ. Technology of Troyes, UMR 6281, CNRS, Troyes, France

⁽²⁾Univ. Lorraine, CRAN, UMR 7039, Campus Sciences, BP 70239, Vandoeuvre-les-Nancy, France

⁽³⁾Univ. Grenoble Alpes, GIPSA-lab, Grenoble, France / CNRS, GIPSA-lab, Grenoble, France

Abstract Condition monitoring (CM) is an essential part in a condition-based maintenance (CBM) program as it can provides useful information about the system health for maintenance decision-making. The quality of CM information depends closely on implemented inspection and instrument technologies, and eventually on invested costs. While numerous works in the literature have considered problems related to CM quality (e.g., imperfect inspection models, detection and localization techniques, etc.), few of them focus on adjusting CM quality for CBM optimization. In this paper, assuming that the quality of CM information can be controlled by adjusting inspection costs, we investigate how such adjustment can contribute to save on the total cost of a CBM program. The study is done for a single-unit deteriorating system subject to random failures and to periodic inspection. We model the quality of the degradation information on the inspected system using the idea that its variance is a decreasing function of inspection cost (i.e., a more accurate CM information requires a more sophisticated inspection, hence a higher cost). A new inspection and replacement strategy whose decisions are adapted to this variance is proposed and compared to a non-adapting classical one. The comparison results under various configurations of system characteristics and maintenance costs allow us to conclude on the interest of CM adjustment in CBM optimization.

1. Introduction

Nowadays, condition-based maintenance (CBM) is widely used in industry because of its interest in improving the durability, reliability, and maintainability of industrial systems (Wang 2008; Jardine et al. 2006). This leads to a steady growth of CBM optimization models in the literature which are more advanced and better adapted to practical industrial concerns. Kong and Park (1997); Huynh et al. (2011); Fouladirad and Grall (2011) investigated the performance of CBM policies with periodic inspections for single-unit stochastic deteriorating systems. On the other hand, different studies aimed to optimize the time interval between two successive inspections have been presented in the literature. In fact, it becomes more interesting to adapt the inspection interval according to the observed level of degradation state (Castanier et al. 2003, 2005) or according to the residual useful life (RUL) of the system (Gebraeel et al. 2005; Yang and Klutke 2001; Do et al. 2015).

However, these above studies are based on the assumption of perfect condition monitoring (CM) which returns the real system states without errors. This assumption seems to be less practical because there might exist some noise and disturbances in CM data, in spite of the progress of censor technology and prognostic techniques. To deal with this problem, numerous works in the literature have been proposed. Newby and Barker (2006) considered an imperfect inspection model in which observed deterioration state is subject to Gaussian error. Ghasemi et al. (2007) developed a partially observed Markov decision process (POMDP) to optimize maintenance policy for the system whose state is hidden and can be estimated based on CM data. Using Hidden Markov Model theory, Neves et al. (2011) addressed the problem of model input estimations from imperfect inspection observations to optimize CBM policies. In summary, these studies attempt to propose efficient approaches to estimate the true degradation state from imperfect data or to develop adequate decision models taken measurement errors into account for maintenance decision-making. But, none of them can reply to the question whether it is necessary to pay higher cost to obtain better accuracy information or not.

In reality, the quality of CM depends closely on implemented inspection and instrument technologies, and eventually on invested costs. Therefore, it could be controlled by adjusting inspection costs, i.e. paying higher cost, to obtain more accurate CM information. The aim of this work is to propose a new hybrid inspection

policy that allows us to adjust CM quality for CBM optimization. The study is based on a single-unit deteriorating system subject to a Gamma deterioration process, (Van Noortwijk 2009), and to periodic inspection, (Kong and Park 1997). The quality of degradation information returned by inspections is characterized by the variance parameter of random errors followed Gaussian distribution, (Newby and Barker 2006). Hence, our contribution is to develop a new flexible maintenance strategy whose decisions are adapted to this variance. It allows us to address the question when the adjustment of inspection quality from low to high level is necessary. Furthermore, comparison of the proposed policy and the non-adapting classical ones are also performed to highlight the interest of CM adjustment in CBM optimization.

This paper is structured as follows: Section 2 presents the the statement and formulation for the problem of adjustment of inspection quality in CBM optimization processes. In section 4, the experimental analysis will be performed in order to examine the performance of proposed policy. Finally, section 5 contains conclusion and prospect of the future work.

2. System modeling and adapted condition-based maintenance policy

Consider a single-unit deteriorating system subject, we assume that the system degradation evolution is assimilated by a homogeneous gamma process, $X(t) \sim G(\lambda \cdot t, \beta)$, widely used in CBM models (Van Noortwijk 2009; Meier-Hirmer et al. 2009; Tan et al. 2010). In the absence of maintenance actions, the system continuously deteriorates from the new state $X(0) = 0$ and the system fails when its deterioration level exceeds a failure threshold L . For a time unit Δt , the probability density of the random increment of the degradation process, ΔX , is given by:

$$f_{\Delta X}(x) = \frac{\beta^{\lambda \Delta t}}{\Gamma(\lambda \Delta t)} x^{\lambda \Delta t} e^{-\beta x} \quad (1)$$

Since the system stops working whenever it fails, one can recognize its failure state without any inspection (i.e., self-announcing failure). However, the system degradation is hidden, monitoring is then required to reveal the degradation level. Continuous monitoring performed at each time step (Δt) is usually very costly, and sometimes impossible in practical engineering applications. In this framework, it is more suitable to implement periodic inspection whose length between two successive inspections (T) is a multiple of Δt . The inspection is instantaneously performed at each discrete inspection time $T_i = i \cdot T$ with a cost c_i^k depending on the inspection quality k (k is the inspection quality index). The system degradation measure at T_i , associated with the quality k , can be expressed by:

$$Z(T_i) = X(T_i) + \varepsilon_k; \quad \varepsilon_k \sim N(0, \sigma_k^2), \quad (2)$$

where σ_k is the variance of errors depending on k . Because of the higher the inspection cost, the more the measure is accurate, σ_k should be an decreasing function of c_i^k .

In the framework of this paper, two type of the inspection quality are considered: low quality ($k = 1$) and high quality ($k = 2$). The inspection quality is characterized by the inspection error variance, σ_k : $\sigma_1 > \sigma_2$ and hence, $c_1 < c_2$. We therefore introduce a hybrid inspection policy, called *IJ* that allow to adjust the inspection quality according to the number of inspections. More precisely, at the i -th inspection time, if $i < J$, low quality inspection is performed ($k = 1$). Otherwise, high quality inspection is implemented ($k = 2$).

Implementing policy *IJ*, the decision to replace or not the system can be directly based on the degradation measure $Z(T_i)$, Newby and Barker (2006):

- If $M \leq Z(T_i) < L$ where M is know as a preventive threshold, the system is preventively replaced by a new one with cost c_p ($c_p > c_i^k$)
- If $Z(T_i) < L$, the system is correctively replaced by a new one with cost c_c ($c_c > c_p > c_i^k$).

Note that a replacement, whether preventive or corrective, can only be instantaneously performed at inspection times. Therefore, there exists a system inactivity after failure, and an additional cost is incurred from the failure time until the next replacement time at a down-time cost rate, c_d . After such replacements, a new renewal cycle starts. The primary objective is to determine the set of decision variables (J, T, M) in order to minimize the

Table 1: The input parameters for numerical examples.

λ	β	L	c_p	c_c	c_d
1	1	12	50	100	30

long run expected cost rate EC_{IJ}^∞ .

$$EC_{IJ}^\infty = \lim_{t \rightarrow \infty} \left[\frac{E[C(t)]}{t} \right] = \frac{E[C(S)]}{E[S]} \quad (3)$$

where $C(t)$ is the cumulative maintenance cost up to time t and S is a renewal cycle.

$$C(t) = c_p \cdot N_p(t) + c_c \cdot N_c(t) + c_d \cdot d(t) + \sum_{k=1}^2 c_i^k \cdot N_i^k(t) \quad (4)$$

$N_i^k(t)$ denoted the inspection number associated with the quality k over time interval $[0, t]$,

$N_p(t)$ and $N_c(t)$ are respectively the accumulated number of preventive and corrective replacements during $[0, t]$,

$d(t)$ is the associated accumulated downtime.

3. Benchmark condition-based maintenance policy

In order to highlight the interest of the proposed inspection policy, we compare its optimal long-run expected cost rate to those of a classical benchmark CBM policy in both the configuration of high and low quality inspection, called CBM_h and CBM_l respectively. The replacement rules of the benchmark policy are similar to our policy but the inspection quality can not be adjusted, then the decision variables are reduced to (T, M) .

Consider the time unit, $\Delta t = 1$, Table 1 presents the input parameters for numerical examples. The characteristics of two considered configurations of the Benchmark policy are defined by:

- **CBM_l policy:** All inspections are performed with low quality, $\sigma_1 = 2$ at low inspection cost, $c_i^1 = 3$.
- **CBM_h policy:** All inspections are performed with perfect quality, $\sigma_2 = 0$ at high inspection cost, $c_i^2 = 3$.

For classic CBM policy, the optimal results describe that:

- **At low quality level:** Inspections are periodically performed after $T = 4$ and the asset is preventively replaced when its estimated degradation state is greater than $M = 7$. The optimal cost rate in this case is $EC_{CBM_l}^\infty = 7.32$

- **At high quality level:** Inspections are periodically performed after $T = 4$ and the asset is preventively replaced when its estimated degradation state is greater than $M = 8$. The optimal cost rate in this case is $EC_{CBM_h}^\infty = 7.26$

We find that with parameter values presented in Table 1, the preventive replacement is postponed when high quality inspections are implemented.

4. Numerical experiments

4.1. Flexibility of Policy IJ

For policy IJ , the optimal cost rate is $EC_{IJ}^\infty = 7.16$. It is less than the results obtained by classic CBM policies (2.24% less than CBM_l policy and 1.47% less than CBM_h policy). The optimal results describe that the system is periodically inspected for every $T = 4$ at low quality level (for the first inspection) and at high quality level (from the second inspection). When the estimated degradation state is higher than $M = 8$, the system is preventively replaced. The flexibility of policy IJ is presented by J . When $J = 1$, policy IJ becomes the high quality configuration of the classical Benchmark policy, CBM_h . When $0 < J < \Theta$, where Θ is a large number, the proposed policy leads to a hybrid policy in which both low quality and high quality inspections are considered. It must be noted that the higher J is, the more interesting the low quality inspections are.

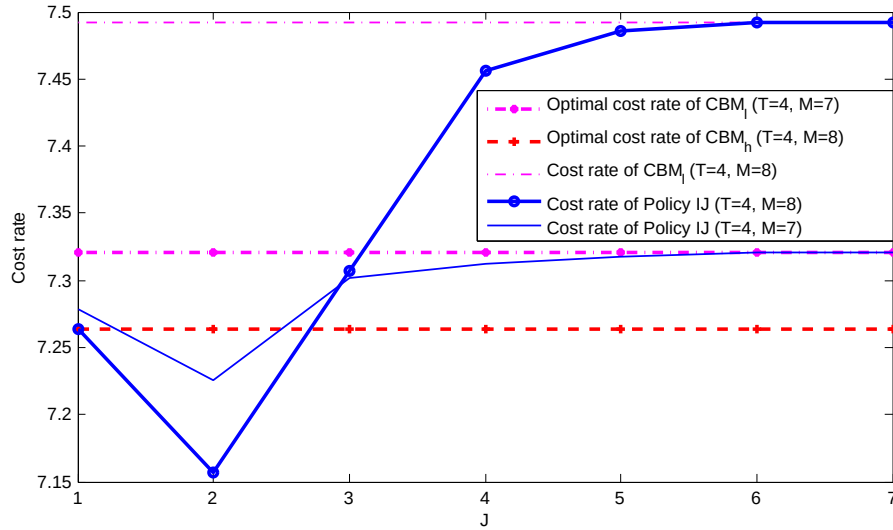


Figure1. Comparison between Policy *IJ* and CBM policies at low and high quality inspection level

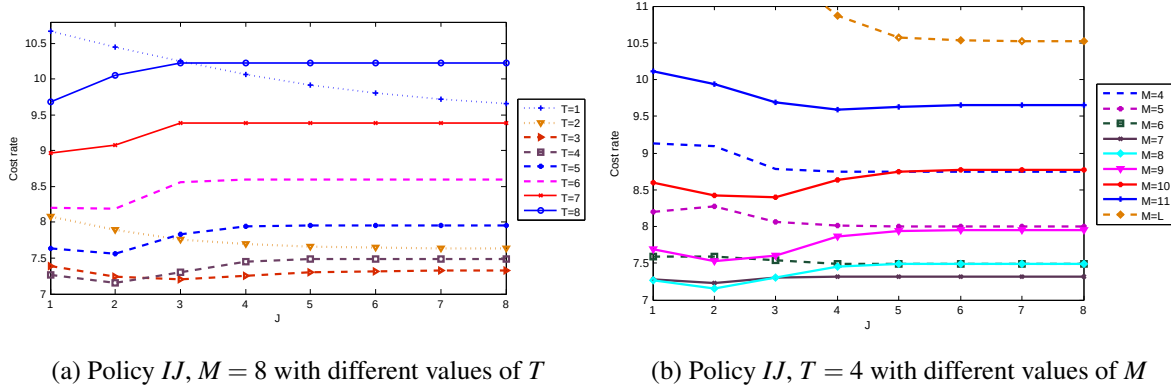


Figure2. Cost rate of Policy *IJ* as a function of *J*

Finally, when $J \rightarrow \infty$, policy *IJ* tends to policy CBM_l , in which all inspections during a cycle of system life are performed at low quality.

The interest of policy *IJ* is illustrated by Figure 1. In this figure, the cost rates of Policy *IJ* (when $(T, M) = (4, 8)$ and $(4, 7)$) follow convex forms having the minimum at $J = 2$. When $J = 1$, policy *IJ* becomes policy CBM_h and if $(T, M) = (4, 8)$, its corresponding cost rate is also the optimal cost rate of policy CBM_h (presented by the red dash-line). When $J \geq 6$, the cost rate of policy *IJ* at $(T, M) = (4, 8)$ converges to the cost rate of policy CBM_l at $(T, M) = (4, 8)$, characterized by the light pink dash-line. Note that the cost rate of policy CBM_l is optimized at $(T, M) = (4, 7)$, presented by the bold pink dash line that is also the convergence value of the cost rate of policy *IJ* at $(T, M) = (4, 7)$ for $J \geq 6$.

Next, we investigate according to every value of T, M , when it is necessary to adjust the inspection quality. Consider at $M = 8$, the dependence of the cost rate of Policy *IJ* on inspection interval T and adjusted threshold J is presented in sub-Figure 2a. We find that the cost rate is :

- non-increasing in J when $T = 1, 2$ (presented by point-dash lines). In order to economize the accumulated inspection cost during a life cycle of the system, it prefers to use low quality inspections for continuous monitoring ($T = 1$) or for short periodic inspection ($T = 2$).
- non-monotone (convex form) in J when $2 \leq T \leq 6$ (presented by dash lines). In these case, it is necessary to adjust the inspection quality from low to high level in order to minimize the maintenance cost rate.
- non-decreasing in J when $T \geq 7$ (presented by continuous lines). For a long periodic monitoring when the inspection point is close to preventive replacement time, it prefers to use high quality inspections.

When $T = 4$, hybrid policy *IJ* could be considered as a classic CBM at low quality inspection level for $J \geq \Theta$, $\Theta = 5$, see sub-Figure 2b, in which all cost rate lines are almost constant for $J \geq 5$. On the other

Table 2: Impact of the low-quality-inspection cost.

c_i^1	1	1.5	2	2.5	3	4	4.5	5
EC_{IJ}^∞	6.5	6.74	6.97	7.11	7.16	7.25	7.26	7.26
T	2	2	3	3	4	4	4	4
M	9	9	8	8	8	8	8	8
J	12	8	4	3	2	2	1	1

hand, we find that the cost rates for $M = 4, 5, 6$ and $M = L = 12$ (presented by dash-lines) are non-increasing in J . That means the low quality inspections are more interest than the high quality inspections. However, for $M = 7, 8, 9, 10, 11$, the cost rates (characterized by continuous lines) follow convex form in J . In these cases, there exists an optimal value J^* such as $1 < J^* < \Theta$, in other words, it prefers to adjust inspection quality from low quality to high quality in order to minimize the maintenance cost rate.

4.2. Impact of the ratio of inspection cost on policy IJ

Consider Table 2, we find that when the low-quality-inspection cost, c_i^1 is increasing (from 1 to the high-quality-inspection cost c_i^2 , $c_i^2 = 5$), the optimal results of policy IJ describe that:

- inspection period T tends to be prolonged,
- preventive replacement threshold M is non-increasing, i.e, the preventive replacement is accelerated,
- the adjustment of inspection quality from low quality to high quality level is accelerated.

In fact, when the ratio between low and high quality inspection cost, noted $\delta_{c_i} = \frac{c_i^2}{c_i^1}$, is high ($\delta_{c_i} \geq 3.3$) regular low quality inspections are enough. For example, $\delta_{c_i} = 5$, the optimal policy describes that $(J, T, M) = (12, 2, 9)$, policy IJ becomes policy CBM_l . The asset is normally preventively replaced before the 12-th inspection ($T_{12} = 24$) when the high quality inspection is implemented for the first time. Otherwise, when δ_{c_i} is low, $1 \leq \delta_{c_i} \leq 1.11$, it prefers to use high quality inspections, then policy IJ becomes policy CBM_h .

5. Conclusion

In this paper, a new CBM policy adopted to the inspection quality is proposed. The adjustment of inspection quality from low to high level is based on the number of inspection J . It allows us to investigate and discuss the interest of a such adjustment of inspection quality in CBM optimization. In fact, the performance of the proposed policy is highlighted through numerous numerical examples. The comparison of the optimal results of the proposed policy is then performed with those of classic CBM at high and low inspection quality level. Thanks to the flexibility, that is presented by decision variable J , the proposed policy gives the better results than other classic one. In details, for several cases, e.g, when the ratio between high and low quality inspection costs is extremely high, the proposed policy tends to a classic CBM at low quality inspection, $J \rightarrow \infty$. Otherwise, the proposed policy might optimally become a classic CBM at high quality inspection, $J = 1$. In general, an adjustment of inspection quality contributes to save on the total cost of a CBM program.

This work focuses on the interest of adjustment of inspection quality in CBM optimization. Therefore, we have simplified the CM-based estimation process of the system's degradation state and used the assumption that the quality inspection is directly characterized by the variance of random errors of the estimated degradation state. It can be seen as a limitation of our model. In further work, the estimation of degradation state based on imperfect inspection information could be taken into account. On the other hands, a more efficient inspection policy that is a residual useful life (RUL) based inspection policy with adjustment of inspection quality could also be investigated.

Acknowledgements

This work has been partially supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01) funded by the French program Investissement d'avenir.

References

- Castanier, B., C. Bérenguer, & A. Grall (2003, October). A sequential condition-based repair/replacement policy with non-periodic inspections for a system subject to continuous wear. *Applied Stochastic Models in Business and Industry* 19(4), 327–347.
- Castanier, B., A. Grall, & C. Bérenguer (2005, January). A condition-based maintenance policy with non-periodic inspections for a two-unit series system. *Reliability Engineering & System Safety* 87(1), 109–120.
- Do, P., A. Voisin, E. Levrat, & B. Iung (2015, January). A proactive condition-based maintenance strategy with both perfect and imperfect maintenance actions. *Reliability Engineering & System Safety* 133, 22–32.
- Fouladirad, M. & A. Grall (2011). Condition-based maintenance for a system subject to a non-homogeneous wear process with a wear rate transition. *Reliability Engineering & System Safety* 96(6), 611–618.
- Gebraeel, N. Z., M. A. Lawley, R. Li, & J. K. Ryan (2005). Residual-life distributions from component degradation signals: A Bayesian approach. *IIE Transactions* 37(6), 543–557.
- Ghasemi, A., S. Yacout, & M. S. Ouali (2007). Optimal condition based maintenance with imperfect information and the proportional hazards model. *International Journal of Production Research* 45(4), 989–1012.
- Huynh, K. T., A. Barros, C. Bérenguer, & I. T. Castro (2011). A periodic inspection and replacement policy for systems subject to competing failure modes due to degradation and traumatic events. *Reliability Engineering & System Safety* 96(4), 497–508.
- Jardine, A. K. S., D. Lin, & D. Banjevic (2006, October). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing* 20(7), 1483–1510.
- Kong, M. B. & K. S. Park (1997, March). Optimal replacement of an item subject to cumulative damage under periodic inspections. *Microelectronics Reliability* 37(3), 467–472.
- Meier-Hirmer, C., G. Riboulet, F. Sourget, & M. Roussignol (2009, September). Maintenance optimization for a system with a gamma deterioration process and intervention delay: Application to track maintenance. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 223(3), 189–198.
- Neves, M. L., L. P. Santiago, & C. A. Maia (2011, October). A condition-based maintenance policy and input parameters estimation for deteriorating systems under periodic inspection. *Computers & Industrial Engineering* 61(3), 503–511.
- Newby, M. J. & C. T. Barker (2006). A bivariate process model for maintenance and inspection planning. *International Journal of Pressure Vessels and Piping* 83(4), 270–275.
- Tan, L., Z. Cheng, B. Guo, & S. Gong (2010). Condition-based maintenance policy for gamma deteriorating systems. *Journal of Systems Engineering and Electronics* 21(1), 57–61.
- Van Noortwijk, J. M. (2009, January). A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety* 94(1), 2–21.
- Wang, W. (2008). Condition-based Maintenance Modelling. In *Complex System Maintenance Handbook*, Springer Series in Reliability Engineering, pp. 111–131. Springer London. DOI: 10.1007/978-1-84800-011-7_5.
- Yang, Y. & G. A. Klutke (2001). A distribution-free lower bound for availability of quantile-based inspection schemes. *IEEE Transactions on Reliability* 50(4), 419–421.

A method of dependent failure-occurrence detection for parallel systems

S. Ota* and M. Kimura**

*Graduate School of Science & Engineering, Hosei University, Tokyo, Japan: shuhei.ota.5b@stu.hosei.ac.jp

**Faculty of Science & Engineering, Hosei University, Tokyo, Japan: kim@hosei.ac.jp

Abstract This paper proposes a statistical test method for 2-unit parallel systems, which discriminates whether the units failed independently or dependently. The method which makes a system redundant by combining several units is often used in order to construct a highly-reliable system. However, if the units break down dependently, the reliability of the parallel system would decrease. We present the statistical test scheme based on the failure occurrence times of the units. It is assumed that the lifetime distribution of the survival unit changes if one unit fails. Therefore our scheme can find dependent failure occurrences by identifying the change of the distribution. The performance of the proposed method is shown by simulation studies.

1. Introduction

In parallel redundant systems, it would be naturally expected that each component fails dependently because of sharing work loads. In such a situation, we are liable to estimate the reliability of the system incorrectly unless we consider the possibility of dependent failure occurrence phenomena. Therefore many researchers have studied the reliability analysis for the dependent failure occurrence as well as the independent one (e.g. Marshall and Olkin (1967)). They made a mathematical model of a 2-unit system which describes the dependent failure environment by using bivariate exponential distribution. In recent years, several studies have applied copula functions to the system models instead of the multivariate exponential distribution. For example, Eryilmaz and Tank (2012) expressed the dependent failure of the 2-unit series system by using FGM copula with exponential marginal distributions. Dobson et al. (2005) constructed another dependent failure model which dealt with cascading failure. The research focuses on a network reliability with a blackout caused by excessive work loads expanding as a snowball. Accordingly, a lot of researchers discussed how they made a model of dependent failure occurrence and how they assessed the system reliability in such a case of the dependent failure occurrence.

In this paper, we investigate the reliability characteristics of the 2-unit parallel redundant system which was proposed in our previous study (see, Ota et al. (2015)) from the following viewpoints; (i) the evaluation of the effect of dependent failure for the lifetime of the system (ii) the development of a methodology which detects a dependent failure occurrence by using components' failure time. In particular, the detection of a dependent failure occurrence contributes to the reliability engineering when we design a highly-reliable system. Our approach which employs the statistical testing scheme for the dependent failure occurrence detection is simple but novel as far as we surveyed in the literature.

First of all, we introduce the dependent failure model of the 2-unit parallel system in Sec. 2. Section 3 explains the algorithm to detect a dependent failure occurrence. Finally, we show the performance of the proposed method through numerical experimentations in Sec. 4.

2. Model description

2.1. Assumptions

In this paper, let us consider a two-unit parallel system and the dependent failure occurrence by the following assumptions.

Assumption I : The parallel system is constructed by two units whose lifetimes are (originally) independent and identically distributed. Figure1 illustrates the system structure.

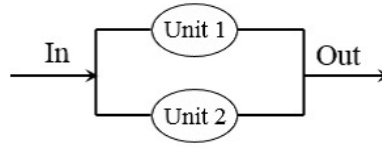


Figure1. 2-unit parallel system.

Assumption II : The hazard rate function of each unit is one of the following: DFR (decreasing failure rate), CFR (constant failure rate) or IFR (increasing failure rate).

Assumption III : If both units survive, the hazard rate function of the units is given by

$$h_1(t) = \lambda_1 m_1 t^{m_1-1} \quad (\lambda_1, m_1 > 0). \quad (1)$$

This assumption describes that the lifetime distribution of the units are Weibull distribution. Let $F(t)$ be the lifetime distribution function of the units, then $F(t)$ can be written by

$$\begin{aligned} F(t) &= 1 - \text{Exp}\left[-\int_0^t h_1(x)dx\right] \\ &= 1 - \text{Exp}[-\lambda_1 m_1 t^{m_1}]. \end{aligned} \quad (2)$$

Assumption IV : If the first failure occurs at t_f , the hazard rate function of the other unit switches to the function $h_2(t|t_f)$ under the condition of t_f . The conditional hazard rate function $h_2(t|t_f)$ is given by

$$h_2(t|t_f) = g(t) - g(t_f) + h_1(t_f) \quad (t > t_f), \quad (3)$$

where

$$g(t) = \lambda_2 m_2 t^{m_2-1} \quad (\lambda_2, m_2 > 0). \quad (4)$$

Assumption V : When the remaining unit fails, the parallel system breaks down.

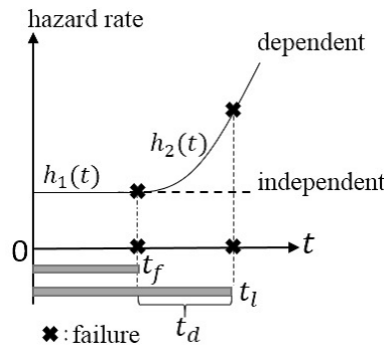


Figure2. Behavior of the hazard rate function.

This model expresses the dependent failure occurrence by changing the hazard rate function. Figure 2 shows an example of hazard rate functions. The lifetimes of the two units obey the distribution function with the hazard rate $h_1(t)$ until the first failure occurs. After the first failure occurrence, the surviving unit obeys $h_2(t)$. In this case, since the hazard rate function switches from CFR to IFR, the surviving unit breaks down sooner than the unit expected originally. Here, we will define the failure occurrence times as follows. Let t_f be the *former failure time* measured from $t = 0$ to the first failure occurrence, and t_l be the *latter failure time* from $t = 0$ to the second failure occurrence. We consider t_d as the *difference time* given by $t_l - t_f$. In addition, we represent T_f , T_d , and T_l as random variables of t_f , t_d and t_l , respectively.

2.2. Lifetime distributions

We can derive the lifetime distributions of the random variables T_f , T_d and T_l . Let T_1 and T_2 are random variables following the hazard rate $h_1(t)$. Then the former failure time T_f is given by $T_f = \min(T_1, T_2)$. Therefore, the

CDF (cumulative distribution function) of T_f , $F_f(t)$, is given by

$$\begin{aligned} F_f(t) &= \Pr[\min(T_1, T_2) \leq t] \\ &= 1 - \text{Exp}[-2\lambda_1 t^{m_1}]. \end{aligned} \quad (5)$$

In addition, its PDF (probability density function) $f_f(t)$ is as follows.

$$\begin{aligned} f_f(t) &= \frac{dF_f(t)}{dt} \\ &= 2\lambda_1 m_1 t^{m_1-1} \text{Exp}[-2\lambda_1 t^{m_1}]. \end{aligned} \quad (6)$$

Let $F_d(t|t_f)$ be the conditional CDF of the difference time T_d . Then $F_d(t|t_f)$ is given by

$$\begin{aligned} F_d(t|t_f) &= 1 - e^{-\int_0^t h_2(x+t_f|t_f)dx} \\ &= 1 - e^{-\lambda_2(t+t_f)^{m_2} + M(t_f)t + \lambda_2 t_f^{m_2}}, \end{aligned} \quad (7)$$

where $M(t_f) = g(t_f) - h_1(t_f)$. The parameters λ_1 and λ_2 are *scale parameters*, m_1 and m_2 are *shape parameters*. In order to get the CDF of the latter failure time T_l , we need to calculate the distribution of the sum of T_f and T_d . Let $F_l(t)$ be the CDF of T_l . Then $F_l(t)$ is derived as follows.

$$\begin{aligned} F_l(t) &= \Pr[T_f + T_d \leq t] \\ &= \int_0^t \Pr[T_d \leq t - t_f | T_f = t_f] f_f(t_f) dt_f \\ &= \int_0^t F_d(t - t_f | t_f) f_f(t_f) dt_f. \end{aligned} \quad (8)$$

Note that $F_l(t)$ does not depend on t_f because t_f is integrated out in Eq. (8).

2.3. MTTF

Considering the MTTF (mean time to failure) is important when we deal with a failure occurrence phenomenon. In this section, we investigate the MTTF of the system comparing with both cases of the dependent failure and independent one. From Eq. (8), the MTTF can be derived by

$$MTTF = \int_0^\infty (1 - F_l(t)) dt. \quad (9)$$

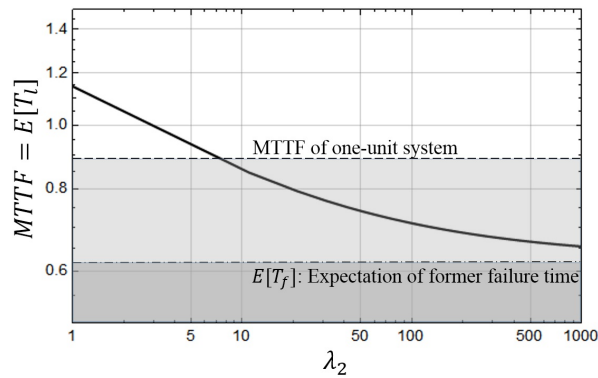


Figure3. Relationship between λ_2 and MTTF. $(\lambda_1, m_1, m_2) = (1, 2, 2)$.

Figure 3 shows that how the MTTF behaves as the parameter λ_2 of the hazard rate function increases. In this figure, parameters λ_1, m_1 and m_2 are fixed as $(\lambda_1, m_1, m_2) = (1, 2, 2)$, and λ_2 is variable. This graph implies that the MTTF of the parallel system decreases as λ_2 increases. Furthermore, MTTF falls below the MTTF of a one-unit system. Hence, we can find that the dependent failure occurrence makes MTTF of the parallel system decline. Therefore, to develop a method which detects the dependent failure occurrence contributes to the improvement of the system reliability.

3. Algorithm of detecting dependent failure

In this section, we explain how the algorithm distinguishes the dependent failure occurrence from the independent one. The algorithm evaluates whether the actual system down time t_l was obtained under the independent failure environment. If an independent failure occurs on the units, T_f and T_l obey $F_f(t) = 1 - e^{-2\lambda_1 t^{m_1}}$ and $F_l(t) = (1 - e^{-\lambda_1 t^{m_1}})^2$, respectively. Note that these functions have the same parameters λ_1 and m_1 . Therefore, if we estimate $\hat{F}_f(t)$ by the method of maximal likelihood from the sample of t_f , then we can estimate $\hat{F}_l(t)$ immediately without using that of t_l . Next, we are able to detect the dependent failure occurrence by performing the goodness-of-fit test between such $\hat{F}_l(t)$ and t_l . As a result of the test, if $\hat{F}_l(t)$ well fits t_l , we can state that the independent failure occurs. On the other hand, if $\hat{F}_l(t)$ does not fit, we can state the dependent failure occurs. We use Kolmogorov-Smirnov test (see, e.g. Feller (1948)) as a goodness-of-fit test. The algorithm to detect a dependent failure occurrence is shown as follows.

Algorithm:

- (i) Estimate $\hat{F}_f(t) = 1 - e^{-2\hat{\lambda}_1 t^{\hat{m}_1}}$ by the method of maximal likelihood with the former failure time t_f .
- (ii) Set $\hat{F}_l(t) = (1 - e^{-\hat{\lambda}_1 t^{\hat{m}_1}})^2$ with the assumption that an independent failure occurs.
- (iii) Calculate Kolmogorov-Smirnov statistics D_n as

$$D_n = \sup_{0 < t < \infty} |\hat{F}_l(t) - S_l(t)|,$$
 where $S_l(t)$ is the empirical distribution function of the latter failure times t_l .
- (iv) Set the null hypothesis H_0 and the alternative hypothesis H_1 as follows.

$$H_0: \text{The latter failure times } T_l \text{ obeys } \hat{F}_l(t).$$

$$H_1: \text{The latter failure times } T_l \text{ do not obey } \hat{F}_l(t).$$
- (v) Perform Kolmogorov Smirnov test with the hypotheses.

If $\sqrt{n}D_n > 1.36$, then reject H_0 with a significance level 5% and accept H_1 . If not, accept H_0 .
- (vi) If H_0 is accepted, an independent failure occurred. If H_1 is accepted, a dependent failure occurred.

Consequently, if the null hypothesis is accepted on the test result for a 2-unit parallel system, then we can state that an independent failure occurred in the system. If not, a dependent failure occurred.

4. Case studies

We investigate how correctly the test method can detect the dependent failure occurrence and independent one. Here, we show the probability of the type I error (denoted by α), and the probability of the type II error (represented by β). In this paper, the type I error is an incorrect statement which means that there was a dependent failure occurrence although there was an independent one actually. Moreover, the type II error is also incorrect statement which declares that there was an independent failure occurrence although there was a truly dependent failure occurrence. $1 - \beta$ is called *power* which means the probability of making a correct judgement when the dependent failure occurred. Therefore, it is desirable that the values of both α and β are small. In our simulation studies, we calculate α and $1 - \beta$ by Monte Carlo simulation (500 times). The data sizes are given by $n = 20, 50, 100$. The significance level is 5%.

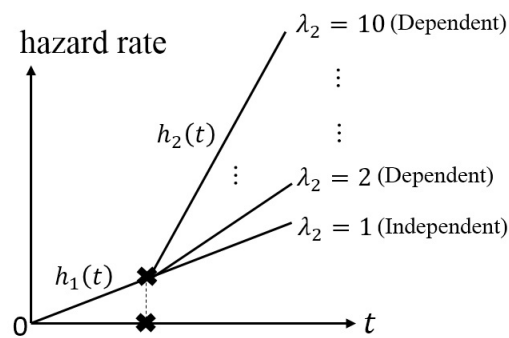
First, we derive the power of the test while only the scale parameter λ_2 is variable. The other parameters are fixed as $(\lambda_1, m_1, m_2) = (1, 2, 2)$, and λ_2 is given by $1, 2, \dots, 10$. Figure 4 illustrates these behaviors of the hazard rate functions. Table 1 shows the power of the test in these ten cases. This result reveals that the power is high as λ_2 and n increase.

We also calculate the power of the test when the shape parameter m_2 is variable. The other parameters are fixed as $(\lambda_1, \lambda_2, m_1) = (1, 1, 2)$, and m_2 is given by $2, 3, \dots, 10$. The power is shown in Tab. 2. Table 2 suggests that the shape parameter m_2 affects the performance of the test as similar as λ_2 .

In summary, the proposed method succeeded in finding the dependent failure occurrence. Especially, it can detect the dependent failure occurrence precisely when data size is large and ratios λ_2/λ_1 and m_2/m_1 are high.

5. Conclusion

This research has discussed the detection method of the dependent failure occurrence phenomenon of the 2-unit parallel system. First, we proposed the stochastic model which describes the dependent failure environment. In this model, lifetimes of the units dependently fail each other by changing the hazard rate functions. Next,

Figure4. Behaviors of the hazard rate functions $(\lambda_1, m_1, m_2) = (1, 2, 2)$.Table 1. Power $(\lambda_1, m_1, m_2) = (1, 2, 2)$.

λ_2	power : $1 - \beta(\%)$		
	$n = 20$	$n = 50$	$n = 100$
1	11.4	13.6	11.2
2	23.0	32.2	58.2
3	39.6	69.2	89.2
4	51.4	85.4	98.6
5	67.6	92.2	99.8
6	68.4	95.8	99.8
7	79.6	98.8	100.0
8	84.6	99.4	100.0
9	84.6	99.8	100.0
10	86.4	99.8	100.0

Table 2. Power $(\lambda_1, \lambda_2, m_1) = (1, 1, 2)$.

m_2	power : $1 - \beta(\%)$		
	$n = 20$	$n = 50$	$n = 100$
2	12.2	10.0	9.4
3	23.8	43.4	64.4
4	43.6	70.6	96.4
5	61.2	91.2	99.8
6	69.8	97.0	100.0
7	80.0	98.6	100.0
8	84.0	99.2	100.0
9	85.2	99.8	100.0
10	86.8	99.8	100.0

we have found that the dependent failure makes the MTTF of the system short. Finally, we have presented the test scheme to find dependent failure occurrence. The advantage of this scheme is to detect the dependent failure occurrence objectively because it is based on the statistical theory. Therefore, the proposed method can contribute to constructing highly-reliable system.

References

- Dobson I, Carreras BA and Newman DE (2005) A loading-dependent model of probabilistic cascading failure. *Probability in the Engineering and Informational Sciences*, Vol. 19, pp. 15-32.
- Eryilmaz S and Tank F (2012) On reliability analysis of a two-dependent-unit series system with a standby unit. *Applied Mathematics and Computation*, Vol. 218, Issue 15, pp. 7792-7797.
- Feller W (1948) On the Kolmogorov-Smirnov limit theorem for empirical distribution. *The Annals of Mathematical Statistics*, Vol. 19, No. 2, pp. 177-189.
- Marshall AW and Olkin I (1967) A multivariate exponential distribution, *J. Amer. Statist. Assoc.*, Vol. 62, pp. 30-44.
- Ota S, Kageyama T and Kimura M (2015) Improvement of reliability evaluation for 2-Unit parallel system with cascading failures by using maximal copula. *IEICE Trans. Fundamentals*, Vol. E98-A, No. 10, pp. 2096-2100.

Acknowledgments

This study is partially supported by JSPS KAKENHI Grant Number 15k01208.

RCM Cost Analysis and Simulation of MV XLPE Cable Failure Process towards Optimum Preventive Maintenance Strategy

M. Qatan* and B.M Alkali*, M. Farrag*, and C. Zhou*,

* Department of Engineering, School of Engineering & Built Environment, Glasgow Caledonian University, G4 0BA, UK ,
mqatan10@gcu.ac.uk, babakalli.alkali@gcu.ac.uk, mohamed.farrag@gcu.ac.uk, c.zhou@gcu.ac.uk

Abstract. In this paper a comprehensive RCM analysis is conducted on the MV XLPE cable failure process to identify all the potential failure modes. An attempt is made to model the individual failure modes contributing to the cable failure process in order to determine cost effective preventive maintenance strategy. A specialist ISOGRAPH RCM cost Availability Workbench Software is used to analyse the reliability of the cable failure and a series cost effective preventive maintenance optimisation graphs are presented to gain a better insight of the failure process of the cables. The results from this study show that cables failures are associated with insulation faults, insulation quality, aging, high temperature, mechanical and electrical stresses, third party damage, poor work quality, and environment factors. The RCM analysis results give a better understanding of the failure process, related failure modes and the cost maintenance as well as the effects on the cables. The results presented in this paper also show the cost contributions of all the individual failure modes of the cable failure. Furthermore in the paper failure consequence analysis of the cables is presented. Results of the safety, operational criticality as well as the safety and operational target of the cables are presented. A discussion of the RCM framework and the assessment of required maintenance practices in order to establish the appropriateness of cost effective and preventive maintenance plans are presented.

Keywords: Reliability Centred Maintenance, preventive maintenance, MV Power Cables, optimisation

1. INTRODUCTION

The cost of effective and planned preventive maintenance activities is important to cable asset manager in order to assure the reliability and availability of cables. Optimal cost effective preventive maintenance policy is expected to improve system reliability and safety performance. The study of adequate maintenance techniques have been emphasized by (Sheriff and Smith 1981) to increased safety and reliability requirements of systems, increased complexity, and rising costs of material and labor. One important research area in reliability is the study of various maintenance policies in order to improve system reliability, to prevent the occurrence of system failure, and to reduce maintenance costs (Pham and Wang 1996). Maintenance, replacement and inspection problems have been widely studied (Sherif 1982). Therefore maintenance and reliability management techniques are needed to support asset manager make decisions on cost effective preventive replacement strategy. Electric power distribution systems constitute a greater risk of interruption to power supply and distribution systems and have received less attention compared to generation and transmission (Bertling 2005). Reliability centred maintenance approach is designed to determine the maintenance requirement of an asset in its operating context and hence identify failures modes that contribute to failures (Moubray 1997).

Maintenance policy on cable asset is expected to be effective if it is conducted just before failure. The cost of preventive replacement generally leads to production losses which have significant impact on the electricity market. The cost of maintenance, replacement of cable assets has been a major challenge over many years and this is as a result of materials used for insulation, insulation faults, insulation quality, aging, high temperature, mechanical and electrical stresses, third party damage, poor work quality, and environment factors. These root cause problem related to cable failures has been one of the major concern raised over the years and it is of main interest in this study.

This paper attempt to link the RCM approach and it applications to cable asset failure data collected from the Sultanate of Oman. MV XLPE cable assets reliability in Oman is vital to the economy as they are used in all of the Oil and Gas production stations, system facilities, electrical substations, and production wells amongst others. Cable failures in electrical power supply normally results in a major loss of production and hence revenue

generated is affected. The MV cables require adequate cost effective maintenance and replacement schedules. In Oman, the current preventive maintenance policy is conducted every 3 years using off line maintenance to determine the health condition of the cable assets.

We investigate the causes of failure and conduct reliability analysis of the cable failure data and predict the future faults, and critical failure modes identified. The cable lifetime distributions are chosen and relevant parameters are selected for the simulation model to determine optimal cost effective preventive maintenance strategy.

2. MEDIUM VOLTAGE (MV) CABLE DESCRIPTION

Medium Voltage cable designs vary widely to meet the diverse requirements of reliability and availability, but there are certain components, which are common to all cables types. All electric cables consist essentially of a low resistance conductor (circular copper or aluminium conductor) to carry the current and insulation (XLPE, paper, or other high resistance insulation) to isolate the conductors from their surroundings. Other main components may include screening to obtain a radial electrostatic field, a metal sheath to keep out moisture or to retain a pressurising medium, armouring for mechanical protection, corrosion protection for the metallic components and a variety of additions extending, for example to internal and external pipes to remove the heat generated in the cable. The cable longitudinal and cross section are shown in Figure 1. The XLPE cable can be applied with a conductor temperature of 130 °C for emergency overload conditions for 10 seconds, it must be understood that erecting and laying conditions necessary to allow for temperatures must be accurately carried out and controlled during installation in order to restrict thermo-mechanical problems.

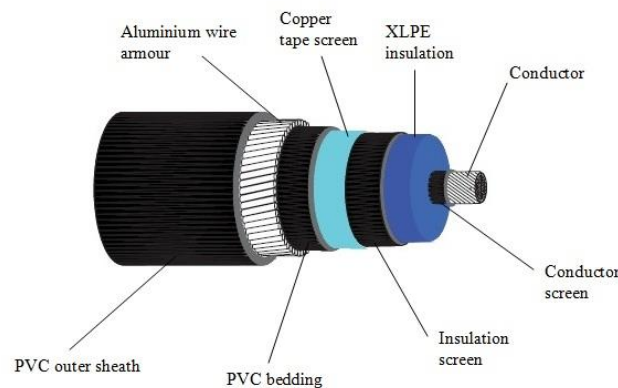


Figure 1: MV cable structure

2.1 Evaluation of cable asset reliability

Reliability assessment of power systems has been developed and evaluated for example see Brown (2002) and Billinton and Allan (1996). Birtwhistle, et al. (2004) investigated the problem of MV cables in North Queensland, Australia and considered the relative economics of replacement. Their study considered failures of MV cables only. Cable and cable termination condition plays a vital role in achieving reliable and constant power supply to operations at all times. Most of the power cable failures occur due to faults in insulation, insulation quality, aging, high temperature, mechanical and electrical stresses, third party damage, poor work quality, and environment factors such as petroleum contaminations, moisture, and dust amongst others.

A study conducted by Bulinski et al. (1992) presents results of cable failure due to water treeing, degradation under combined mechanical and electrical stresses in cable insulation. They also consider aging and high temperature as critical factors failure in a cable failure process. Shwebdil et al., (2003) conducted a comprehensive study of thermal aging on medium voltage cable insulation materials in Dhahran, Saudi Arabia. Their result from Shwebdil et al., (2003) shows that the cable insulation will not operate at 90° C.

However for temperature above 107°C, XLPE gets soften somewhat and continues to provide good insulation at temperatures at up to about 150° C. Birtwhistle, et al. (2004) addressed the aging problem due to

high voltage stress leading to high temperatures which cause the cable age acceleration. The aging and degradation as a result of temperature has been highlighted by Dissado and Fothergill (1992). The temperature is a main problem related to cable failures and have raised concern over the condition of many of cables over the years and temperature is of interest in this study. The average temperature in Oman is between 24 – 56° C and the operating temperature of the cable and cable termination is average between 70 – 104° C. Cable condition analysis is complex and many related factors affect its performance. MV cables require adequate monitoring in order to schedule cost effective planned maintenance and replacement of defective units.

3. RELIABILITY CENTERED MAINTENANCE

The RCM approach originated from the airline industry, a report by Nolan and Heap (1978) gives an in-depth assessment of maintenance function in the airline industry. In this paper, RCM analysis has been conducted to identify the failure modes and the maintenance requirements of the cables assets. It is observed that the cable failures occur due to fault in insulation, third party damage and environmental factors such petroleum contaminations, moisture, and dust amongst others. The outcome of the RCM is analysed further using the RCMCost Availability Workbench Software. The RCMCost module of Availability Workbench (AWB) provides the full framework for building the RCM model to represent the system as shown in Table 1. The list of failure modes associated with the cable insulation and conductor is presented in Table 1 and 2 respectively.

Table 1: Cable insulation failure modes

ID	Failure mode	ID	Failure mode
B.1.1.1	Fault in insulation	B.1.3.2	Conductor damage
B.1.1.2	Circuit with earth fault	B.1.3.3	Sheath damage
B.1.1.3	Third-party damage	B.1.3.4	Circuit with earth fault
B.1.1.4	Electrical stress	B.1.3.5	Leakage to earth
B.1.1.5	Mechanical stress	B.1.3.6	Third-party damage
B.1.1.6	Water tree	B.1.3.7	Electrical stress
B.1.1.7	Environment factors	B.1.3.8	Mechanical stress
B.1.1.8	Improper installation	B.1.3.9	Environment factors
B.1.2.1	Fault in insulation	B.1.3.10	Improper installation
B.1.2.2	Third-party damage	B.1.4.1	Third-party damage
B.1.2.3	Source control error	B.1.4.2	Environment factors
B.1.2.4	External impulse	B.1.4.3	Improper installation
B.1.2.5	Electrical stress	B.1.5.1	Improper installation
B.1.2.6	Mechanical stress	B.1.5.2	Use wrong tools
B.1.2.7	Water tree	B.1.5.3	Third-party damage
B.1.2.8	Environment factors	B.1.5.4	Electrical stress
B.1.3.1	Fault in insulation		

Table 2: Cable conductor failure modes

ID	Failure mode	ID	Failure mode
A1.1.1	Open circuit	A.1.2.6	Environment factors
A.1.1.2	Circuit with earth fault	A.1.2.7	Improper installation
A.1.1.3	Loss connection	A.1.3.1	Conductor deteriorated
A.1.1.4	Wrong design selection	A.1.3.2	Conductor failure
A.1.1.5	Improper installation	A.1.3.3	Insulation deteriorated
A.1.2.1	Loss connection	A.1.3.4	Fault in insulation
A.1.2.2	Bad compression	A.1.3.5	Source control error
A.1.2.3	Wrong size	A.1.3.6	External impulse
A.1.2.4	Third-party damage	A.1.3.7	Electrical stress
A.1.2.5	Electrical stress	A.1.3.8	Environment factors

For convenience of failure mode identification in our analysis using the software, we used default ID to represent the failure modes as shown in Table 1 and 2. The cost contribution of the failure modes (causes) is assessed for the insulation and conductor, the costs associated to the failure modes (causes) is presented in the graphs in Figure 2a and 2b respectively.

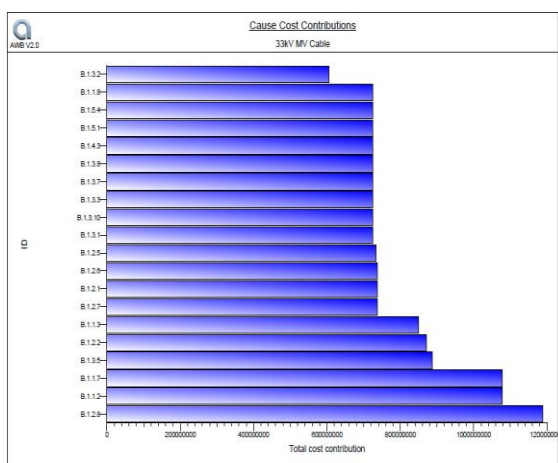


Figure 2a. Cost contribution for insulation

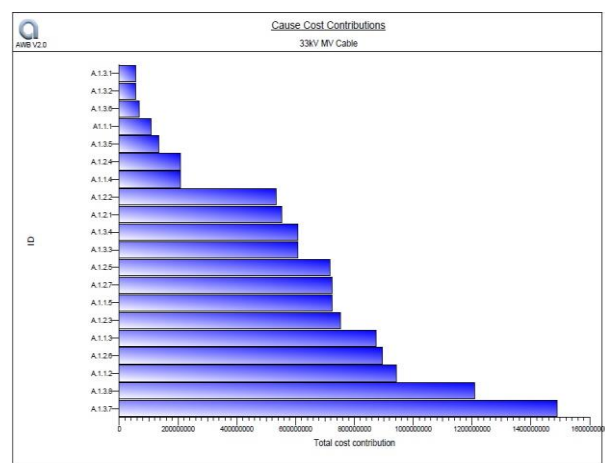


Figure 2b. Cost contribution for conductor

Notice from the graphs the environment factors (ID B.1.2.8) of the insulation and the electrical stress (ID A.1.3.7) of the conductor are the highest contributors to the failure of cables. The results however are not surprising as this is what we expect considering the system. But the cost contribution is staggering. We focus on the 7 most significant contributors (i.e for insulation as ranked are environment factors, fault in insulation, improper installation, electrical stress, conductor damage, and circuit with earth fault) and (for conductor ranked are the electrical stress, wrong size, improper installation, environment factors, circuit with earth fault, fault in insulation, and loss connection. There are many models that have been proposed by researchers for example on the application of RCM on cables has been conducted and the data and knowledge gained from the survey has been put into an RCM framework as an application study for a cable system component. For example (Bertling et al., 2002; Bertling et al., 2003).

3.1 The RCMCost Module Simulation

Simulation methods are generally employed in reliability studies when deterministic methods are incapable of modeling strong dependencies between failures. In addition simulation can readily handle the reliability behaviour of components with constant and non-constant failure or repair rates. Monte Carlo Simulation is used to emulate the chance variations that affect the cable failure. To do this we used the historical data and input the data into the program and run simulation to generate 1000 random numbers that form a uniform distribution.

3.2 ISOGRAPH (AWB) Model Simulation Results

The lifecycle of the cable form information obtained from the company is 20 years. However, there is not justification on how the lifetime of the cable asset is determined. We use the age of the cable as input parameter in our simulation model. The cable asset status is considered as working or in a failed state, if in a failed they cannot be repaired, the maintenance scheme conducted on the cable is replacement which assumes a renewal process in case failure of the cable has an exponential distribution.

The frequency of failure and the margin of operational losses as well as cost associated to the cable conductor and insulation allowed is specified in the model. The program assigns criticality ranking automatically. The highest 7 ranked failure mode (cause) cost prediction simulation result of the cable insulation show that the effect are due to corrective and scheduled maintenance cost of failures to the cable. The result also show that the total cost contributions for each failure mode is the cost due to the use of labor, equipment, operational, and spares as shown in Table 3. Similarly the cause cost prediction contribution result for the conductor is presented in Table 4.

Table 3: Failure mode cost simulation result for insulation

SN	ID	Failure mode	Equipment cost contribution (\$)	Labor cost contribution (\$)	Operational cost contribution (\$)	Effect cost contribution (\$)	Spare cost contribution (\$)	Total cost contribution (\$)
1	B.1.2.8	Environment factors	2732137.98	2428601.81	1132727.77	1920820346	3830387.77	1930944200.96
2	B.1.3.1	Fault in insulation	3985877.77	3311082.27	937820.94	1125991457	1892528.87	1136118766.83
3	B.1.3.10	Improper installation	3985877.77	3311082.27	937820.94	1125991457	1892528.87	1136118766.83
4	B.1.3.7	Electrical stress	3985877.77	3311082.27	937820.94	1125991457	1892528.87	1136118766.83
5	B.1.3.8	Mechanical stress	3985877.77	3311082.27	937820.94	1125991457	1892528.87	1136118766.83
6	B.1.3.2	Conductor damage	1396834.35	1240961.32	581122.69	978049065	1984680.05	983252663.44
7	B.1.3.4	Circuit with earth fault	1396834.35	1240961.32	581122.69	978049065	1984680.05	983252663.44

Table 4: Failure mode cost simulation result for the conductor

SN	ID	Failure mode	Equipment cost contribution (\$)	Labor cost contribution (\$)	Operational cost contribution (\$)	Effect cost contribution (\$)	Spare cost contribution (\$)	Total cost contribution (\$)
1	A.1.3.7	Electrical stress	3979213.42	3305165.17	934267.13	1172963368.09	1873144.32	1183055158.12
2	A.1.2.3	Wrong size	3979213.42	3306125.56	934267.13	1170224182.91	1873144.32	1180316933.33
3	A.1.1.5	Improper installation	3979213.42	2951060.47	934267.13	1125840058.12	1873144.32	1135577743.45
4	A.1.3.8	Environment factors	865757.25	560698.83	361856.79	612992396.39	1203246.52	615983955.77
5	A.1.1.2	Circuit with earth fault	865757.25	768567.77	361856.79	606892446.95	1252242.16	610140870.92
6	A.1.3.4	Fault in insulation	2613412.88	2171064.26	634655.45	601716411.03	1254594.88	608390138.49
7	A.1.1.3	Loss connection	363239.75	368329.20	299225.48	569627542.89	1256417.29	571914754.62

From the ranked cause cost predictions contributions the common cause cost contributors from the insulation and conductor of the cable are environment factors, electrical stress, and fault in insulation, improper installation, as well as circuit with earth fault. Also other contributors are critical causes to the cable due to the failure modes effect. The criticality simulation determined the environmental, operational, and safety consequences directly associated with the cause contributions. The criticality values are represented the severity of the effects associated with the cause combined with its frequency of occurrence. The results of the lifetime criticality of the failure modes with safety, operational and environmental contributions related to the insulation and the conductor as shown in the Table 4 and 5 respectively.

Table 4: Lifetime criticality for insulation

Safety criticality	0.07255
Operational criticality	4.37700
Environmental criticality	0.01287

Table 5: Lifetime criticality for the conductor

Safety criticality	0.02844
Operational criticality	3.00700
Environmental criticality	0.00490

The RCM criticality conducted using the software programme is used to estimate criticality ranking and total cost contribution of the failure modes (causes). The range of the criticality ranking is between 0-1 and notice that estimate safety, environmental and the outage frequency is very negligible. The operational criticality estimates for the insulation have range between the values of 0.09 to 0.3 and on the other hand the estimate for the conductor has range between 0.03 and 0.3. The effect of the operational criticality failure mode is associated with lost production costs as a consequence. The analysis highlighted the important contributors to operational costs and safety and environmental risks of the insulation and conductor as shown in Table 6 and 7.

Table 6: RCM criticality analysis estimates for the insulation

SN	Cause ID	Failure mode	Safety criticality	Operational criticality	Environmental criticality	Mean outage frequency	Total cost contribution (\$)
1	B.1.2.8	Environment factors	0.005682	0.188500	0.001008	0.000367	1930944200.96
2	B.1.3.1	Fault in insulation	0.002755	0.269700	0.000489	0.000181	1136118766.83
3	B.1.3.10	Improper installation	0.002755	0.269700	0.000489	0.000181	1136118766.83
4	B.1.3.7	Electrical stress	0.002755	0.269700	0.000489	0.000181	1136118766.83
5	B.1.3.8	Mechanical stress	0.002755	0.269700	0.000489	0.000181	1136118766.83
6	B.1.3.2	Conductor damage	0.002893	0.095980	0.000513	0.000190	983252663.44
7	B.1.3.4	Circuit with earth fault	0.002893	0.095980	0.000513	0.000190	983252663.44

Table 7: RCM criticality analysis estimates for the conductor

SN	Cause ID	Failure mode	Safety criticality	Operational criticality	Environmental criticality	Mean outage frequency	Total cost contribution (\$)
1	A.1.3.7	Electrical stress	0.002755	0.314400	0.000489	0.000179	1183055158.12
2	A.1.2.3	Wrong size	0.002755	0.308400	0.000489	0.000179	1180316933.33
3	A.1.1.5	Improper installation	0.002755	0.269700	0.000489	0.000179	1135577743.45
4	A.1.3.8	Environment factors	0.001783	0.068810	0.000316	0.000120	615983955.77
5	A.1.1.2	Circuit with earth fault	0.001783	0.063290	0.000316	0.000120	610140870.92
6	A.1.3.4	Fault in insulation	0.001788	0.073740	0.000317	0.000154	608390138.49
7	A.1.1.3	Loss connection	0.001781	0.029490	0.000316	0.000155	571914754.62

4. PREVENTIVE MAINTENANCE OPTIMISATION

This section present the preventive maintenance optimisation of the cable failure causes in terms of the expected cost per unit time. The effect of using PM strategy is analysed and an estimate of cost effective PM strategy is proposed. The simulation model is display the cost effective solutions. The graphs presented are considered as the highest contributors of the insulation and conductor. The PM optimisation simulation result in Figure 3 for the environment factors of the insulation give optimum PM cost recommendation at point 8760 operating hour.

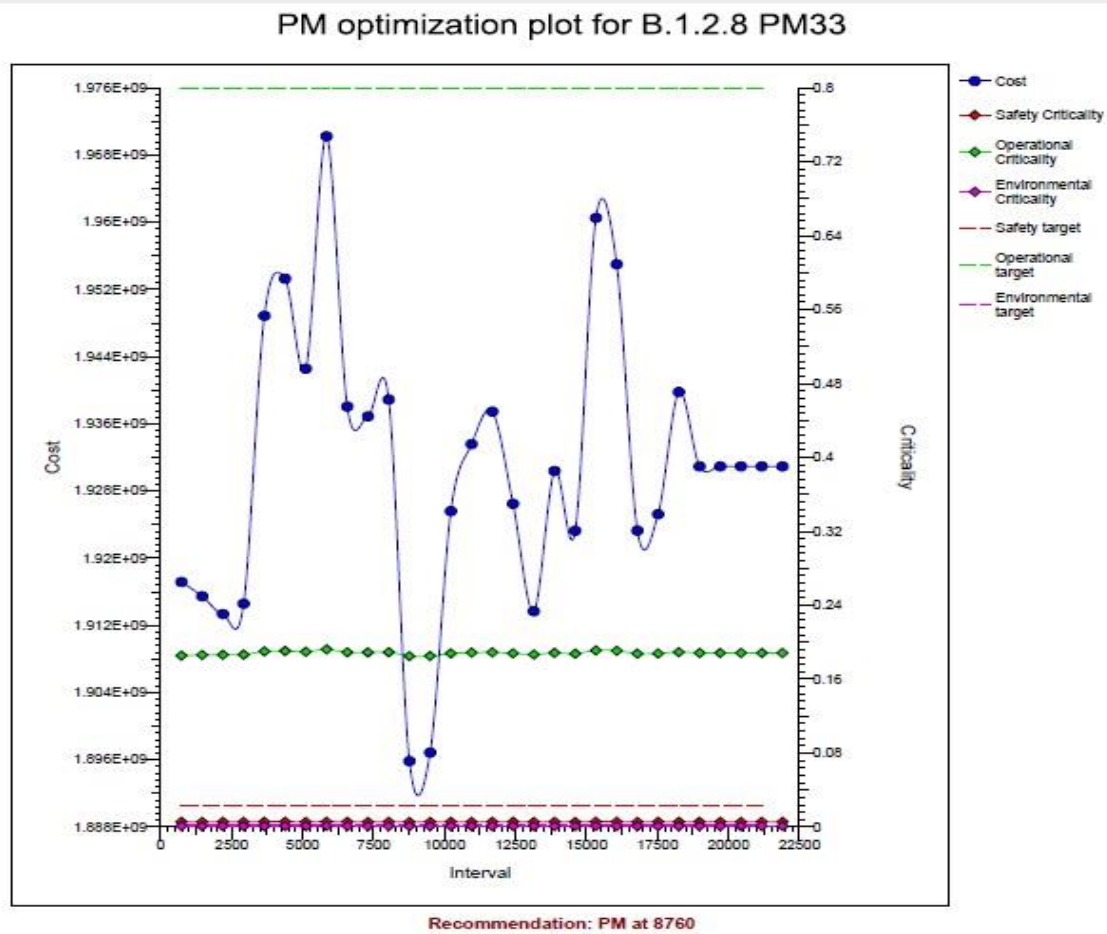


Figure 3. The PM optimisation plot for the environment factors of the insulation

Notice after 8760 operating hour the cost curve is not constant with time, it is getting constant from 18500 operating hours. The assessment of the environmental, operational, and safety criticality for the environment factors related failures and their severity as depicted on the graph is constant.

Similarly, the assessment result of PM optimisation for the electrical stress for the conductor related failures. The optimum PM cost recommendation at point 13140 operating hour. One would be able to make the decision on preventive replacement of the cables assets in practice from the result obtained in this study.

The data assessed from the complete population of cable over 20 year aging period. In practice PM is expected to be conducted during the lifetime of the cable assets. However the number of occasions and timing of PM should depend on the cost of performing the PM and the cost benefit of doing so.

5. CONCLUSION

This paper presents a comprehensive RCM analysis conducted on MV XLPE cable asset and by using Availability Workbench software PM optimisation is also conducted. The results from the application show that reliability centred maintenance approach is used to establish link the RCM approach and reliability analysis and compare cost effective PM intervals. The study also shows that RCM can be performed and the resulting output can be used as real data input within a simulation framework to optimise a PM strategy. Furthermore, the recommendations from this study give asset manager a very good insight on the cable asset failure pattern and the outcome of study can be used to support decisions on preventive replacement, age replacement and block replacement policy.

ACKNOWLEDGMENTS

The author would like to express their gratitude to Petroleum Development Oman for their support on providing data information.

REFERENCES

- Bertling L, Eriksson R, Allan R, Gustafsson L, Ahlen M (2002). *Survey of Causes of Failures Based on Statistics and Practice for Improvements of Preventive Maintenance Plans*, PSCC, Sevilla 2002.
- Bertling L, Eriksson R, Allan R (2003). *Impact of Maintenance Strategy on Cost and Reliability of Distribution Systems*, International Conference on Electricity Distribution, Barcelona 2003.
- Bertling L, Allan R, and Erikson R (2005). A Reliability Centered Asset Maintenance Method For Assessing the Impact of Maintenance in Power Distribution Systems. *IEEE Transactions on Power Systems*. Vol. 20(1) pp. 75-82. 0885-8950.
- Billinton R and Allan R (1996). *Reliability Evaluation of Power Systems*. 2nd ed. New York: Plenum
- Birtwhistle D, Lyall J S, Footitt F, Wickramasuriya P, Gilbert R, Powell L, and Saha T (2004). *An Accelerated Wet Ageing Test on Medium Voltage XLPE Cables*, Australasian Universities Power Engineering Conference (AUPEC 2004).
- Brown ER (2002). *Electric Power Distribution Reliability*. New York: Marcel Dekker
- Bulinski A, Bamji S S, Braun J M, and Densley J (1992). *Water treeing degradation under combined mechanical and electrical stresses, presented at Electrical Insulation and Dielectric Phenomena*, 1992. Annual Report., Conference on, 1992.
- Dissado L A and Fothergill (1992). *Electrical Degradation and Breakdown in Polymers*, Peter Peregrinus Ltd, London, IEE, ISBN:0-863-41196-7.
- Moubray J (1997). *Reliability Centered Maintenance*. 2nd edition. New York: Industrial Press Inc. ISBN 0831131462, 9780831131463.
- Nolan F S and Heap H F (1978). *Reliability Centered Maintenance*. Springfield. VA National Technical Information Service, US Department of Commerce.
- Pham H, Wang HZ (1996) Imperfect maintenance. *European Journal of Operational Research* 94:425-438.
- Sherif YS, Smith ML (1981) Optimal maintenance models for systems subject to failure - A review. *Naval Research Logistics Quarterly* 28/1:47-74.
- Sherif Y.S., "Reliability analyses. Optimal inspection and maintenance schedules for failing systems", *Microelectronics Reliability* 22 (1982) 59-115.
- Shwebdil M H, Mors M A, Abugurain A (2003). *Thermal aging tests on XLPE and PVC cable insulation materials of Saudi Arabia*, Conference on Electrical Insulation and Dielectric Phenomena, IEEE 2003.

Modelling of operational reliability and maintenance cost for avionics systems with permanent and intermittent failures

A. Raza* and V. Ulansky**

*DOPA, Maintenance Group, P.O. Box: 372, Abu Dhabi, UAE, ahmed_awan@hotmail.com

**National Aviation University, 1 Kosmonavta Komarova Avenue, Kiev 03058, Ukraine, vulanskyi@yahoo.com

Abstract. Current aviation data suggest that there are more than 400,000 no fault found (NFF) cases per year. The existence of the NFF phenomenon has a negative impact upon avionics systems safety, availability and life cycle cost. Intermittent failures are the leading causes of NFF events in avionics systems. In this study, we examine the process of operation and maintenance of the line replaceable unit (LRU) in the presence of permanent and intermittent failures with an arbitrary distribution of time to failure. It is assumed that the LRU is continuously tested and both types of failures are automatically detected by the built-in test equipment (BITE). A mathematical model is developed to calculate the LRU operational reliability, the probability of the LRU recovery and the cost of maintenance. Different types of avionics systems redundancy are considered. A numerical example illustrates the evaluation of operational reliability and maintenance costs for a modern avionics system.

1. Introduction

Nowadays, Boeing and Airbus aircraft families use digital avionics. Modern avionics are often shared functionally by specific blocks, which are also called LRUs. Avionics systems are redundant and comprise several easily removable LRUs. Each LRU has its own BITE, which provides continuous testing. Any avionics LRU operates to a safe failure (permanent or intermittent), which is registered during flight or after landing. LRUs rejected by BITEs are dismantled and then can be re-tested and repaired. Thus, the safety of modern aircraft is provided by the use of redundant avionics systems, while the regularity of flights is ensured through the use of a sufficient number of spare LRUs. The described maintenance strategy can be ineffective in the case of a high rate of intermittent failures. Unfortunately, conventional test equipment do not confirm the presence of intermittency in the dismantled LRUs and the same failures can happen again in the next flights.

According to ARINC 672 (2008), a NFF is the result of testing when a unit removed as faulty at one level of maintenance is found to be fault free when tested at the next lower level of maintenance. Steadman et al. (2008) pointed out that avionics problems accounted for 75% of all NFF cases in the aviation industry. The NFF rates for avionics systems is between 20% and 50% (Söderholm, 2007). The impact of NFF on the aircraft fleet includes an increase in maintenance and repair time, reduction of the flight regularity, as well as an increase in spare LRUs, which ultimately leads to an increase in the cost of avionics lifecycle. Thus, it is necessary to consider the impact of intermittent failures on the operational reliability and cost of avionics maintenance.

Consider some references related to the operational reliability assessment of systems subject to intermittent failures. Prasad (1992) analysed a Markovian reliability model of digital systems subject to both permanent and intermittent faults. The system under operation can be in one of three states. Kranitis et al. (2006) considered a reliability model of optimal periodic testing of intermittent faults in an embedded pipelined processor. A two-state continuous-parameter Markov model is used for modelling the intermittent faults. Hsu Y. and Hsu C. (1991) analysed a three state Markov model for fault-tolerant systems by taking the effects of permanent and intermittent faults into consideration. Nakagava (2008) considered a communication system subject to intermittent faults. It is supposed that faults have an exponential distribution and are hidden. Faults become permanent failures when the duration in hidden state exceeds an upper limit time. Nakagava (2005) considered an inspection policy to intermittent faults. The testing of a system is planned at regular intervals to detect the faults. Exponential distribution of time to a permanent and intermittent failure is assumed. Raza and Ulansky (2015) considered a mathematical reliability model of a continuously tested LRU subject to permanent and

intermittent failures. Mathematical expressions for average availability of redundant systems are derived considering the spare part system sufficiency. Schen et al. (2016) considered the stochastic Petri net based formalization model of intermittent fault. The parameters for an intermittent fault are computed.

2. Problem statement

The effectiveness of some avionics systems is largely determined by the given probability of the system to perform its functions during flight. An indicator of the effectiveness of systems to ensure flight safety can be the probability of failure-free operation, which we define as operational reliability. A second measure of the maintenance effectiveness can be the cost of maintenance required to meet the given operational reliability. At first we consider the case when the avionics system is a single-unit, i.e., it includes only one LRU. Let us denote Ξ ($\Xi \geq 0$) as a random time to LRU permanent failure with probability density function (PDF) $\omega(\xi)$. The LRU condition is continuously monitored by its BITE and in the case of a permanent failure the LRU is switched off. In the case of an intermittent failure the LRU is usually not switched off during flight. However, information about the intermittent failures that occurred during flight is recorded by the on-board computer.

The following procedure is set for recovery operations. If a permanent or intermittent failure occurred in flight, after landing at the base airport the LRU is dismantled and directed for repairing. We assume that after any repairing the LRU becomes as good as new.

Further, under the operational reliability we mean the probability of LRU failure-free operation on the interval (t_k, t) considering the fact that at time points $t_1, \dots, t_k$ the unscheduled repairing of the LRU with permanent or/and intermittent failures were conducted. It is required to determine the operational reliability and the probability of LRU recovery with permanent and intermittent failure at an arbitrary and exponential failure time distribution. Then these indicators, as well as the cost of maintenance will be determined for redundant avionics systems.

3. Single-unit avionics systems

We introduce the following notations: $B(t_j)$ is the event consisting in restoration of the LRU at time t_j after j -th flight; $h_{IF}(t_j)$ and $h_{PF}(t_j)$ are the events consisting in restoration of the LRU with intermittent and permanent failure, respectively; $P_R(t_j)$, $P_{IF}(t_j)$ and $P_{PF}(t_j)$ are, respectively, the probabilities of the events $B(t_j)$, $h_{IF}(t_j)$ and $h_{PF}(t_j)$; $R(t_k, t)$ is the LRU operational reliability in the interval (t_k, t) , where $t_k < t \leq t_{k+1}$.

To determine $R(t_k, t)$, $P_R(t_j)$, $P_{IF}(t_j)$, and $P_{PF}(t_j)$, we introduce the PDF of random variables $\Xi, \Theta_1, \dots, \Theta_j$, which we denote as $\Omega(\xi, \theta_1, \dots, \theta_j)$, where Θ is the time to intermittent failure with PDF $\psi(\theta)$ and

$$\Theta_j = \Theta - \sum_{v=1}^{j-1} (t_v - t_{v-1}) \quad (1)$$

is the remainder of the operating time to intermittent failure after $j - 1$ flights ($j = 1, 2, \dots$), where $t_v - t_{v-1}$ is the duration of the v -th flight and $t_0 = 0$. By the multiplication theorem of the PDFs, we can write

$$\Omega(\xi, \theta_1, \dots, \theta_j) = \omega(\xi) \Omega_0(\theta_1, \dots, \theta_j | \xi), \quad (2)$$

where $\Omega_0(\theta_1, \dots, \theta_j | \xi)$ is the conditional PDF of random variables $\Theta_1, \dots, \Theta_j$ under the condition that $\Xi = \xi$.

We introduce two conditional probabilities associated with intermittent failures. The conditional probability of occurrence of an intermittent failure in the j -th ($j = 1, 2, \dots$) flight is formulated as follows:

$$P_{IF|O}(\overline{t_1, t_{j-1}}; t_j | \xi) = P \left[\bigcap_{i=1}^{j-1} \Theta_i > (t_i - t_{i-1}) \cap \Theta_j < (t_j - t_{j-1}) | \xi \right], \quad \xi > t_j. \quad (3)$$

The conditional probability of not occurring an intermittent failure during the j -th flight is stated as follows:

$$P_{\overline{IF}|O}(\overline{t_1, t_{j-1}}; t_j | \xi) = P\left[\bigcap_{i=1}^j \Theta_i > (t_i - t_{i-1}) | \xi\right].$$

(4)

The probabilities (3) and (4) are determined by integrating the conditional PDF $\Omega_0(\theta_1, \dots, \theta_j | \xi)$ over the corresponding limits:

$$P_{IF|O}(\overline{t_1, t_{j-1}}; t_j | \xi) = \int_{t_1-t_0}^{\infty} \int_{t_2-t_1}^{\infty} \dots \int_{t_{j-1}-t_{j-2}}^{\infty} \int_0^{t_j-t_{j-1}} \Omega_0(u_1, \dots, u_j | \xi) du_1 \dots du_j,$$

(5)

$$P_{\overline{IF}|O}(\overline{t_1, t_{j-1}}; t_j | \xi) = \int_{t_1-t_0}^{\infty} \int_{t_2-t_1}^{\infty} \dots \int_{t_j-t_{j-1}}^{\infty} \Omega_0(u_1, \dots, u_j | \xi) du_1 \dots du_j.$$

(6)

Theorem 1. The following relations hold:

$$R(t_k, t) = \sum_{j=0}^k P_R(t_j) \int_{t-t_j}^{\infty} P_{\overline{IF}|O}(\overline{t_{j+1}-t_j, t_{k-1}-t_j, t_k-t_j} | \mathcal{G}) \omega(\mathcal{G}) d\mathcal{G}, \quad t \geq t_k,$$

(7)

$$P_R(t_j) = P_{IF}(t_j) + P_{PF}(t_j),$$

(8)

$$P_{IF}(t_j) = \sum_{v=0}^{j-1} P_R(t_v) \int_{t_j-t_v}^{\infty} P_{IF|O}(\overline{t_{v+1}-t_v, t_{j-1}-t_v, t_j-t_v} | \mathcal{G}) \omega(\mathcal{G}) d\mathcal{G},$$

(9)

$$P_{PR}(t_j) = 1 - \sum_{v=0}^{j-1} P_R(t_v) \int_{t_j-t_v}^{\infty} P_{\overline{IF}|O}(\overline{t_{v+1}-t_v, t_{j-2}-t_v, t_{j-1}-t_v} | \mathcal{G}) \omega(\mathcal{G}) d\mathcal{G}.$$

(10)

Proof. Let us begin with the proof of (8). The event $h_{IF}(t_j)$ will be the case, if an intermittent failure occurs in the j -th flight. The event $h_{PF}(t_j)$ will occur, if a permanent failure occurs in the j -th flight. The system will be restored at time point t_j , if there will be any of the events $h_{IF}(t_j)$ or $h_{PF}(t_j)$. Therefore,

$$B(t_j) = h_{IF}(t_j) + h_{PF}(t_j).$$

(11)

Assuming that $h_{IF}(t_j)$ and $h_{PF}(t_j)$ are mutually exclusive events and applying to (11) the addition theorem of probability we obtain relation (8).

For the proof of (7), (9) and (10) we write the probability definition of indicators $R(t_k, t)$, $P_{IF}(t_j)$ and $P_{PF}(t_j)$.

The operational reliability $R(t_k, t)$ is represented as

$$R(t_k, t) = P\left\{\bigcup_{j=0}^k \left[B(t_j) \cap \Xi > t - t_j \cap \left(\bigcap_{i=j+1}^k \Theta_i > t_i - t_{i-1} \right) \right] \right\}.$$

(12)

Then we formulate the probabilities

$$P_{IF}(t_j) = P\left\{\bigcup_{v=0}^{j-1} \left[B(t_v) \cap (E_1(t_j - t_v) \setminus E_2(t_j - t_v)) \right] \right\},$$

(13)

$$P_{PF}(t_j) = 1 - P\left\{\bigcup_{v=0}^{j-1} [B(t_v) \cap E_1(t_j - t_v)]\right\}, \quad (14)$$

where

$$E_1(t_j - t_v) = \Xi > t_j - t_v \cap \left(\bigcap_{i=v+1}^{j-1} \Theta_i > t_i - t_{i-1}\right)$$

is the event consisting in the fact that the LRU, which begins to work at time t_v , does not fail for a time $t_j - t_v$, and no intermittent failure occurs during the flights $v + 1, \dots, j - 1$;

$$E_2(t_j - t_v) = \Xi > t_j - t_v \cap \left(\bigcap_{i=v+1}^j \Theta_i > t_i - t_{i-1}\right)$$

is the event different from $E_1(t_j - t_v)$ only in the fact that at the j -th flight also no intermittent failure occurs; $\setminus$ is the symbol of the two events difference.

Assuming $P_R(t_0) = 1$ and applying to expressions (12) - (14) the addition and multiplication theorems of probability we obtain (7), (9) and (10). *Q.E.D.*

As shown by Drenick (1960), the exponential distribution is an appropriate distribution of permanent failures for complex systems. According to Qin et al. (2005), a flight director system may consist of 460 digital integrated circuits (ICs), 97 linear ICs, 34 memories, 25 ASICs, and 7 processors. Such systems have a large number of electronic components. For these components, external failure mechanisms and intrinsic failure mechanisms can cause the components to fail. Different failure modes come together to form a constant LRU failure rate. Therefore, exponential distribution of time to permanent failure is appropriate for most modern avionics LRUs

$$\omega(t) = \lambda e^{-\lambda t}, \quad (15)$$

where λ is the LRU permanent failure rate.

Let us assume that intermittent failures are also subject to an exponential distribution

$$\psi(t) = \theta e^{-\theta t}, \quad (16)$$

where θ is the LRU intermittent failure rate.

Since the exponential distribution has memoryless property, (5) and (6) are converted to the following form:

$$P_{IF|O}(\overline{t_1, t_{j-1}}; t_j | \xi) = \exp\left[-\theta \sum_{i=1}^{j-1} (t_i - t_{i-1})\right] \left\{1 - \exp\left[-\theta(t_j - t_{j-1})\right]\right\}, \quad (17)$$

$$P_{\overline{IF}|O}(\overline{t_1, t_{j-1}}; t_j | \xi) = \exp\left[-\theta \sum_{i=1}^j (t_i - t_{i-1})\right]. \quad (18)$$

Corollary 1. If the LRU has an exponential distribution of time to permanent and intermittent failure, then the following relations hold:

$$R(t_k, t) = \sum_{j=0}^k \left\{ P_R(t_j) \exp\left[-\theta \sum_{i=j+1}^k (t_i - t_{i-1}) - \lambda(t - t_j)\right] \right\}, \quad (19)$$

$$P_{IF}(t_j) = \sum_{v=0}^{j-1} P_R(t_v) \exp\left[-\theta \sum_{i=v+1}^{j-1} (t_i - t_{i-1}) - \lambda(t_j - t_v)\right] \left\{1 - \exp\left[-\theta(t_j - t_{j-1})\right]\right\}, \quad (20)$$

$$P_{PF}(t_j) = 1 - \sum_{v=0}^{j-1} P_R(t_v) \exp\left[-\theta \sum_{i=v+1}^{j-1} (t_i - t_{i-1}) - \lambda(t_j - t_v)\right]. \quad (21)$$

Proof. Substituting (15), (17) and (18) in (7), (9) and (10) we obtain (19), (20) and (21). *Q.E.D.*

Corollary 2. If the LRU has an exponential distribution of time to permanent and intermittent failure and the flight duration is constant and equal τ , i.e., $t_j = j\tau$, then the following relations hold:

$$R(k\tau, t) = \sum_{j=0}^k P_R(j\tau) \exp\left\{-\left[(k-j)\theta\tau + \lambda(t-j\tau)\right]\right\}, \quad (22)$$

$$P_{IF}(j\tau) = (1 - e^{-\theta\tau}) \sum_{v=0}^{j-1} P_R(v\tau) \exp\left\{-\left[(j-v-1)\theta + (j-v)\lambda\right]\tau\right\}, \quad (23)$$

$$P_{PF}(j\tau) = 1 - \sum_{v=0}^{j-1} P_R(v\tau) \exp\left\{-\left[(j-v-1)\theta + (j-v)\lambda\right]\tau\right\}. \quad (24)$$

Proof. Substituting $t_j = j\tau$ into (19)-(21), we obtain (22)-(24). *Q.E.D.*

Example 1. Assuming in (22) that $t = (k+1)\tau$ calculate (22)-(24) if $\lambda = 10^{-4} \text{ h}^{-1}$, $\theta = 2 \times 10^{-4} \text{ h}^{-1}$ and $\tau = 5 \text{ h}$.

The results of calculations are shown in Table 1. As can be seen from Table 1, starting from the fourth flight all probabilities reached steady-state values. It should be also noted that P_{IF} is two times greater than P_{PF} .

Table 1. Calculated probabilities depending on the number of flight

k	$P_R(k\tau)$	$P_{IF}(k\tau)$	$P_{PF}(k\tau)$	$R[k\tau, (k+1)\tau]$
0	1,0	0	1,0	0.999500124979169
1	0.001498875562289	0.000999000541458	0.000499875020831	0.999498252229370
2	0.001498876161360	0.000999001140529	0.000499875020831	0.999500125577941
3	0.001498874964116	0.000999000542057	0.000499874422059	0.999500124979169
4	0.001498875562289	0.000999000541458	0.000499875020831	0.999500124979169
5	0.001498875562289	0.000999000541458	0.000499875020831	0.999500124979169

Since the random process of changing the LRU states is regenerative, then by virtue of the limit theorem for regenerative processes the following limits must exist:

$$\begin{cases} \lim_{k \rightarrow \infty} R[k\tau, (k+1)\tau] = R^*(\tau), \lim_{j \rightarrow \infty} P_{IF}(j\tau) = P_{IF}^*(\tau), \\ \lim_{j \rightarrow \infty} P_{PF}(j\tau) = P_{PF}^*(\tau), \lim_{j \rightarrow \infty} P_R(j\tau) = P_R^*(\tau). \end{cases} \quad (25)$$

The proof of the limits existence is omitted because of its bulkiness. An indirect proof are the data in Table 1.

The limits (25) make it possible to calculate the steady-state values of a number of the effectiveness indicators.

Number of removals due to intermittent failures during time T (total hours flown) is given by

$$N_{IF}(T) = P_{IF}^*(\tau) T / \tau. \quad (26)$$

Number of removals due to permanent failures during time T is

$$N_{PF}(T) = P_{PF}^*(\tau) T / \tau. \quad (27)$$

Total number of removals during time T is

$$N_R(T) = P_R^*(\tau) T / \tau = N_{IF}(T) + N_{PF}(T). \quad (28)$$

One of the most important indicators of avionics maintenance effectiveness is the mean time between unscheduled removals (MTBUR). As stated by Kayton and Friend (1997), for many avionics systems MTBUR is about 50% of MTBF. This results in 40 ÷ 50% increase in direct operating costs (Hess, 2015).

Theorem 2. The following formula holds for MTBUR:

$$MTBUR = \tau / P_R^*(\tau). \quad (29)$$

Proof. According to IATA (2015), MTBUR is defined as follows:

$$MTBUR = \text{Total hours flown} / \text{Total unscheduled removals}. \quad (30)$$

Substituting T and $N_R(T)$ from (28) to (30) we obtain (29). *Q.E.D.*

Example 2. Calculate MTBUR using data of Example 1.

Substituting $\tau = 5$ h and $P_R^*(\tau) \approx 0.0015$ to (29) gives $MTBUR = 3336$ h, which is only 33.4% of MTBF.

4. Redundant avionics systems

Consider the case of active redundancy when all m identical LRUs are in operational state during flight.

For a parallel structure of avionics system the steady-state value of operational reliability is given by

$$R_m^*(\tau) = 1 - [1 - R^*(\tau)]^m. \quad (31)$$

For an h -out-of- m structure of avionics system the steady-state value of operational reliability is

$$R_m^*(\tau) = \sum_{i=h}^m \binom{m}{i} [R^*(\tau)]^i [1 - R^*(\tau)]^{m-i}. \quad (32)$$

Equations (31) and (32) can easily be proved by the method of structure functions (Barlow and Proshan, 1981).

As a cost indicator we use the total maintenance cost of a redundant avionics system during time T

$$C(T) = \frac{mT}{\tau} [C_{IF} P_{IF}^*(\tau) + C_{PF} P_{PF}^*(\tau)], \quad (33)$$

where C_{IF} and C_{PF} are, respectively, the mean cost of repairing LRU with an intermittent and permanent failure.

Example 3. Modern wide-body aircraft usually incorporate tripled instrument landing system (ILS), i.e., $m = 3$. The ILS influences on the flight safety. Therefore, operational reliability, MTBUR and total cost of maintenance can be considered as the key indicators of maintenance effectiveness. Assume that $T = 4000$ h, $\lambda = 10^{-4} \text{ h}^{-1}$, $\tau = 5$ h, $C_{IF} = 1000$ £, and $C_{PF} = 2000$ £. Calculation of $R^*(\tau)$, $MTBUR$ and $C(T)$ as a function of θ .

The steady-state value of $R^*(\tau)$ is independent of θ ; therefore, operational unreliability, $1 - R^*(\tau)$, is 1.25×10^{-10} for any θ . Figure 1 (a) and (b) show the dependence of MTBUR and $C(T)$ on θ . As can be seen from Fig. 1(a), MTBUR of a single ILS is very much affected by θ . From Fig. 1(b) follows that $C(T)$ begins to rise significantly when $\theta > 10^{-4} \text{ h}^{-1}$.

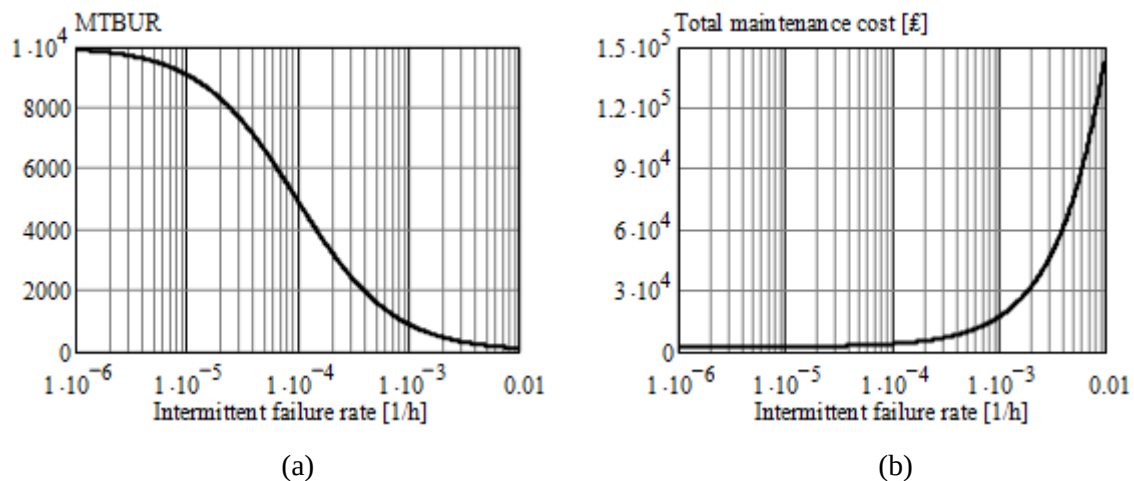


Figure 1. Dependence of MTBUR (a) and total maintenance cost (b) on intermittent failure rate.

5. Conclusions

In this study, we have derived mathematical equations for assessing operational reliability and maintenance cost of single-unit and redundant avionics systems subject to intermittent and permanent failures. The proposed mathematical model allows determining some very important indicators of avionics maintenance effectiveness such as the number of removals due to the intermittent and permanent failures during any time of operation, MTBUR and total maintenance cost. Numerical analysis has shown that high rate of intermittent failures significantly reduces MTBUR and increases the total maintenance cost, which is confirmed by statistical data from various sources.

References

- ARINC Report 672 (2008) Guidelines for the reduction of No Fault Found (NFF). *Avionics Maintenance Conference*, Aeronautical Radio Inc.
- Steadman B, Berghout F, Olsen N, Sorensen B (2008) Intermittent fault detection and isolation system. In *Proc. IEEE Autotestcon*. Salt Lake City, UT, 8-11 Sept. IEEE, pp.37-40.
- Söderholm P (2007) A system view of the no fault found phenomenon. *Rel. Eng. and System Safety*, 92(1), 1-14.
- Prasad VB (1992) Digital systems with intermittent faults and Markovian models. In *Proc. of 1992 IEEE 35th Midwest Symp on Circuits and Systems*. Washington, DC, 9-12 Aug. IEEE, v. 1, pp. 195-198.
- Kranitis N, Merentitis A, and Laoutaris N (2006) Optimal periodic testing of intermittent faults in embedded pipeline processor applications. In *Proc. of 2006 IEEE Design, Automation and Test in Europe Conf.* Munich, Germany, 6-10 March. IEEE, v. 1, pp. 1-6.
- Hsu Y and Hsu C (1991) Novel model of intermittent faults for reliability and safety measures in long-life computer systems. *International Journal of Electronics*, 71(6), pp. 917-937.
- Nakagava T (2008) *Advanced reliability models and maintenance policies*. Springer, 234 p.
- Nakagava T (2005) *Maintenance theory of reliability*. Springer, 269 p.
- Raza A and Ulansky V (2015) Minimizing total lifecycle expected costs of digital avionics' maintenance. In *Proc. of the 4th TESConf2015*. Cranfield, UK, Nov. 3-4. Procedia CIRP, v. 38, pp. 118-123.
- Shen Q, Qiu J, Lin G, and Lv K (2016) Intermittent fault's parameter framework and stochastic Petri net based formalization model. *Maintenance and Reliability*, 18(2), pp. 210-217.
- Drenick R (1960) The failure law of complex equipment. *J. of the Soc. for Ind. and Appl. Math.*, 8, pp. 680-690.
- Kayton M and Friend WR (1997) *Avionics navigation systems*. John Wiley & Sons, Inc.
- Hess R (2015) Electromagnetic environment. In *Digital avionics handbook* (Spitzer SR et al., Eds.), CRC Press.
- IATA (2015) Guidance material and best practices for inventory management.
- Barlow RE and Proshan F (1981) *Statistical Theory of Reliability and Hypothesis Testing*. McArldle Press, Inc.

Research on the application of association rules in fault diagnosis of high-speed railway in China

Ren Jun-Hua* and Liu Feng** Hu Hui***

*Beijing Jiaotong University, Beijing, China, renjunhua@bjtu.edu.cn

**Beijing Jiaotong University, Beijing, China, fliu@bjtu.edu.cn

*** Beijing Jiaotong University, Beijing, China, huhui@bjtu.edu.cn

Abstract: With the rapid development of China's high-speed railway network and the increment of operating mileage, the requirement for the fault diagnosis and safety of high-speed railway has been increasing significantly. In order to enhance the safety and reduce the failure rates, it is essential to analyze the factors that closely related to safety and failures of high-speed railway. Compared with the traditional failure diagnosis method, the association rule algorithm of data mining technology has the advantage of dealing with large scale fault data efficiently. In the present study, the parallel apriori algorithm based on MapReduce algorithm was proposed. Furthermore, a subsystem of association analysis regarding fault diagnosis, which was based on mining association rules and integrated with the parallel association rules algorithm, was constructed by taking advantage of the sparsity characteristic of the fault data. The structure and function of the established subsystem was validated by test data. The proposed fault diagnosis subsystem and the distributed parallel association rules can be made use of in the fault diagnosis of China's high-speed railway.

1. Introduction

From beginning of Beijing-Tianjin intercity high-speed rail operation in August 2008, the high-speed railway network in China has possessed the largest scale and highest speed around the world. The total operating mileage of high-speed railway in China reaches 19,000 kilometers. Furthermore, the four lengthwise high-speed railway network has been constructed. There are more than 8,000 motor train unit in China and the total running bullet train is more than 2,100 pairs. The high-speed railway has become the main vehicle type of people transportation. Under the circumstance, the high-speed railway plays a key role in public security, including national economy security and people's lives. Therefore, the fault diagnosis and security management of high-speed railway is essential and crucial.

Association rule is one of the most important data mining methods aiming at extracting unknown and potential useful rules from large amount of data. Agrawal R, Imieliński T and Swami A(1993) firstly proposed the association rule and the Apriori algorithm based on frequent itemsets. The Apriori algorithm is based on recursive statistics, and the frequent itemset is justified sheared by minimal support degree. The other algorithms are mainly based on Apriori algorithm, or its variants and extensions. Due to the process of multiple original data scanning and generation of a large number of candidate itemsets, the Apriori algorithm has the disadvantages of low efficiency in data mining and large space of usage memory. With regards to the shortcomings, many researchers have proposed many optimized methods. Park J S, Chen M S and Yu P S(1995) proposed the Data Handling Program (DHP) algorithm, which uses the hashing technique to efficiently improve the generation process of candidate sets. Ashok S, Edward O and Shamkant(1995) proposed the Partition algorithm based on the division strategy on the purpose of reducing the number of scanning database in the data mining process and the CPU burden. Hannu T(1996) obtained some worthy rules based on sampling optimization method by extracting samples from transaction database. Then the remainder of database was used to verify the algorithm. The method was verified to significantly reduce the cost of input and output. These algorithms have improved the efficiency of data mining association rules in different degrees. Nevertheless, the inherent defects in Apriori algorithm has not been overcome completely. In practice, the data used to be mined are unnecessary to be added, reduced and modified. On the other hand, the threshold value of minimum support and minimum confidence will also be modified in order to achieve the interesting association rules. Under the circumstances, the initially obtained association rules may not be a strong rule and a new association rule may be obtained with refreshed data. Therefore, the association rules need to be updated. The traditional approach is to use the updated data set or the new threshold value to do the data mining

once again, whereas the initial result of data mining is wasted. Therefore, the traditional method is low-efficiency and undesirable. The advisable method is to obtain the desired association rules by efficiently use the initial data mining results. At present, there are some scholars studying incremental update mining of association rules, and some related algorithms have been proposed by Bi Jianxin and Zhang Qishan(2005). Nevertheless, with the increase of database, the efficiency of these algorithms needs to be further improved.

The integrated management platform of high-speed railway is suitable to be established by lightweight Enterprise Service Bus(ESB) technology. From the perspective of the overall solution, Intel Corporation provides the Hadoop software as the solution through operating and optimizing the software platform. The advantage of Intel Hadoop is its computing framework with highly parallel and scalable distributed batch operation, whose processing ability is close to real-time mode. At the same time, the high performance distributed massive data warehouses Hive is a data warehouse architecture built on Hadoop to support the querying and analyzing of massive data stored in Hadoop.

With the rapid construction and development of high-speed railway in China, the number of the EMU and maintenance organization grows fastly. At the same time, a large amount of data are accumulated in the operation and maintenance of high-speed railway. These data has the characteristics of wide distribution, huge quantity, heterogeneous, and small effective information ratio. In addition, in order to meet the requirement of high efficiency and accuracy of EMU operation and maintenance, it needs to quickly and accurately extract the relevant information to form a variety of services information and support its operation and maintenance. Therefore, the data of high-speed railway EMU operation and maintenance has obvious characteristics of big data, which also brings a great challenge to EMU operation and maintenance. At present, the multi-source and heterogeneous data analysis and fault correlation diagnosis of China's high-speed railway EMU and its key components based on data analysis just begin. There is a large gap for practical usage. In the present study, a fault diagnosis method and its application in high-speed railway EMU in China is investigated based on association rules of big data mining.

2.Parallel improvement of the Apriori algorithm

2.1 Definition of multiple minimum support association rule

The mining strategy of multiple minimum support proposed Wang Zhenyu, Shi Lei Bai and Xiong Fanguan(2002) is a data mining method, which defines the value of minimum support association rule respectively according to the characteristics of transaction itself based on the defined support degree .

Definition 1 For a data item set $I = \{i_1, i_2, \dots, i_m\}$, its any transaction is given the minimum support degree that it needs to satisfy, which is called the minimum support degree, and is denoted as MIS .

Definiton 2 For mining rule: $X_1 \wedge X_2 \wedge X_3 \wedge \dots \wedge X_k \Rightarrow X_{k+1} \wedge \dots \wedge X_n, X_i \in I(i=1,2,\dots,k,k+1,\dots,n)$, the minimum value of minimum support of the various items contained in the rule $\min(MIS(X_1), MIS(X_2), \dots, MIS(X_n))$ is the minimum support degree required to be satisfied.

2.2 The Apriori algorithm of multiple minimum support degree based on MapReduce

The event items with high occurrence frequency will have a higher minimum support degree restriction and vice versa. Under the circumstances, the minimum support required for any candidate rule will be generated in the premise of the minimum support degree for each item in the rule itself. Therefore, the algorithm is designed as follows.

Input: data set DB, minimum support degree set MIS

Output: frequent item set L

$C_1^* = \text{database } D$

$L_1 = \{c | c \in C_1^*, c.count / |DB| \geq MIS(c)\}$

for($k = 2; L_{k-1} \neq \emptyset; k++$)

if($k=2$) then $C_2^* = L_2 - \text{candidate} - \text{gen}$

```

else  $C_k^* = \text{candidate} - \text{gen}(L_{k-1})$ 
end
for each transaction  $t \in DB$  do
  for each candidate  $c \in C_k^*$ 
    Map(): if(c is in t) c.count++
  end
end
Reduce():  $L_k = \{c \mid c \in C_k, c.\text{count} / |DB| \geq \text{MIS}(c[1])\}$ 

 $L = \cup L_k$ 
end

```

In the MapReduce programming, the Apriori algorithm makes calculation in the form of key value pair (Key/Value), and calculation results are exported in the form of key value pair (Key/Value) too. Map (image) and Reduce (statute) are the most important two methods in MapReduce processing.

2.2.1 The definition of MapReduce input and output

Defining the format of input and output Key/Value is the first step in implementation of MapReduce programming function. The definition of key value pair is shown in Table 1. Furthermore, the Key/Value pair can be used to realize the Map function and the Reduce function.

Table 1
The input and output definitions of Map and Reduce functions

input/output	Map function	Reduce function
input: key/value pair	Key: Line number Value: data in one row	Key: candidate set Value:1
output: key/value pair	Key: candidate set Value:1	Key: Frequent set Value: support degree

The MapReduce system in cloud computing environment divides the original data into many small data within the size of defined input slice. Furthermore, with the function of load balancing, the MapReduce system can put the data slice in each node of the cloud computing cluster evenly, used as the input for the Map task. Then, each node will process the small data of input slice using the distributed parallel computing. The Apriori parallel algorithm based on MapReduce function solve the problems of data distribution and node communication effectively. The input of Map function takes the line position of affairs as the key and the transactional context as value. The output of Map function uses the project in transaction or candidate itemsets as key and number 1 as value. The Reduce function statutes these key/value pairs and the numbers 1 corresponding to each key is accumulated. Then, the final output is set as the supporting number obtained from the accumulation. The projects or supports of candidate item set takes as final output. The parallel algorithm of Apriori uses the distributed parallel computing environment of cloud computing and realizes the parallel scanning of data set.

The Map function mainly collects the appearance frequency number of each item in candidate set. The Reduce function mainly prunes the candidate sets containing non-frequent itemsets. Every frequent set is produced by the implementation of Map function and Reduce function. The data set is divided into small pieces transformed to each datanode after the data is stored in HDFS. The Map function executes the data block, thereby generating a <key, value> pairs for each recorder . The Hadoop framework gathers automatically the <key, value> pairs together in accordance with the same items. Then, it transfers the candidate itemsets value list to the Reduce function. The Reduce function accumulates these values and generates a statistical value as a synchronization candidate set. The advantage of the algorithm is that there is no data exchange between the data nodes. What transferred between the data nodes is the statistical values. In each traversal of the data set, each Map function generates a local candidate set, then the Reduce function collects these local statistics data together, forming the whole candidate set. Figure 1 shows the phases of the data flow and the parallel Apriori algorithm in the MapReduce programming model. The multiple iterations of the MapReduce calculation forms the total data analysis. Each iteration generates a frequent set till a closed frequent set is formed.

Map:(row_id, trasaction) →list(itemset, v2=1)
 Reduce: (itemset, list(v2)) →(itemset, sum(v2))

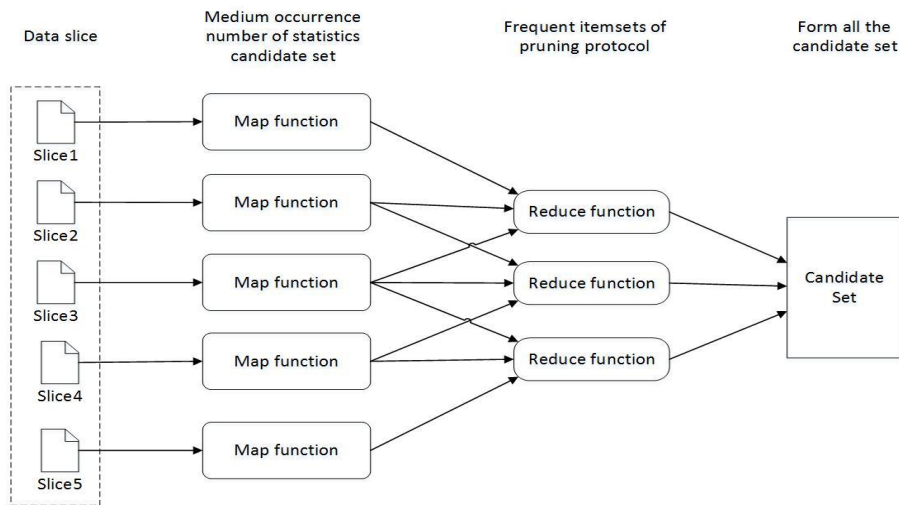


Figure1. The phases of parallel Apriori algorithm programming model based on MapReduce

2.2.2 The Map function Design

In order to carry out rapid information transmission in each computing node in the Hadoop cluster, it is needed to serialize these information into binary data. The Hadoop provides its own serialization Writables to match the platform. The program written in MapReduce and run on Hadoop is used to serialize the input key/value pair. Two methods are defined in the Writable interface. One method is used to write the information into the binary format as DataInput stream. The other method is used to read the information from the binary format of DataOutput.

A new extension of Writables interface Frequentitemsets class is redefined in the MapReduce, which can be used to storage each generated frequent itemsets to FrequentItemsets class, and then the candidate item sets can be obtained through the algorithm Apriori_gen (L_k). Finally, the numbers in each item in candidate item sets can be achieved through scanning the local data.

The code is designed as follows.

```

Map(key,value,  $L_k$ , x.count, context)
{
   $C_k = L_k$  Links to generate maximum frequent item sets
  for(i=1, i<=  $C_k$ .count; i++)
  {
    if (  $C_k$ .contains(value)==true )
    {
       $C_k$ .item The corresponding record is 1.
      Context.write(  $C_k$ .item, 1);
    }
  }
}
  
```

2.2.3 Reduce function Design

The main function of Reduce is to put the results of the output of the Map function into the data in the Reduce function, and add the same item count, judging whether they are greater or equal to the setting support degree. Finally, it outputs the satisfied results.

The Reduce function code is designed as follows.

```

Reduce(key,value, context)
  
```

```

{
    IntWritable r=new IntWritable();
    int count=0;
    for(IntWritable iw:value)
    {
        Count+=iw.get(); // Counting the same items in the candidate set
    }
    if(count> Individual support for each item MIS(i))
    {
        r.set(sum);
        context.write(key,r);
    }
}

```

3.Case study

A lot of data is generated in the process of EMU's operation and maintenance, these data are collected by the "wireless transmission equipment (WTD)" installed in the EMU, then send to the ground server used wireless network after organizing and processing. The running state of EMU can be effectively monitor and managed through data analysis after the ground receiver receiving. The data used in this case is the information recorded of CRH2 type EMU all parts of China in 2014, the average daily failure record is 1.5 millions, and the total size of the data is about 500G. Data files are derived from the CSV Oracle format file, using Hive to complete the data cleaning work and making HQL (SQL class) statement into the MapReduce program. The EMU fault mining experimental procedure of MapReduce algorithm based on Apriori is as follows:

I .The EMU fault data is imported into the Hive based Hadoop cluster, and stored in the distributed file system HDFS of Hadoop cluster, Setting file input and output path, the system reads the data slice (split) from the HDFS and transmits it to each map node;

II . The minimum support and minimum confidence are set before data mining, then the Hadoop cluster is started, and frequent item sets of fault data are mined. The calculated output of the map node become the input of the reduce node, the mining results are stored in the specified output directory after the calculation completed;

III. Using MR Apriori algorithm for mining, the frequent items calculated are stored in HDFS;

Table 2

The partial results from association rules mining

Fault name	Rule	Support count	Degree of support	Confidence level
Abnormal air valve	CRH2,6018A,05	27812	1.85%	9.51%
Air condition and ventilation system	CRH2,2109A, October	25398	1.69%	11.37%
Traction motor	CRH2,6036L, 0.9 million km	27350	1.82%	7.64%
Oil flow of Traction transformer rises	CRH2,6017L,310km/h	18941	1.26%	12.62%
Buzzer tweeting in lavatory	Waste discharge valve abnormal 280km/h	37469	2.49%	13.7%
braking system	CRH2,6037L,M1,13 million km	15398	1.03%	9.32%
Abnormal air valve, air condition and ventilation system	ongitude:110.14-122.43, latitude: 22.5-42.37	37569	2.51%	10.26%
The efficiency of regenerative braking systems reduces	CRH2,2136A,MOkm/h, 1.5 million km	24719	1.64%	8.41%
voltage abnormality of AC auxiliary power supply system	CRH2, the current of DC auxiliary power supply system rise	17491	1.17%	8.2%

IV .The program running process and cluster state are viewed through HDFS management page <http://localhost:50070/>, monitoring the health status of Name node and Data node. The mining results can be obtained from HDFS after the program runs.

V .The strong fault association rules of the EMu are generated using frequent item sets, and the support and confidence of each rule are calculated.

The experimental data is cited from Hu Hui's Research and Implementation of Mining Association Rules for EMU Failure Data Based on Hadoop. On the basis of the above experimental procedure, the association rules of fault data set of EMU are mined using the improved association rule algorithm, the results of the frequent and related faults are obtained from the Hadoop cluster mining and analyzed. Some association rules are shown in table 2.

For example, the association rule "CRH22109A, October=>air condition and ventilation system[support=1.69%, conf=11.37%]" means the fault of air condition and ventilation system frequently happens on train 2109A in CRH2 every October, which degree of support is 1.69%, and the confidence level is 11.37%. The fault association rule can give some suggestions for EMU maintenance, i.e. the air condition and ventilation system of train 2109A in CRH2 should be maintained in October each year in order to exclude it before fault happens.

For the association rule "longitude: 110.14-122.43 latitude: 27.5-42.37 => anomal air valve, air condition ventilation system [support=2.51%, conf= 10.26%]" means the fault on new air valve abnormal and ventilation and air condition system frequent happens in the area longitude 110.140 - 122.430 and latitude: 22.50-42.370, The degree of support is 2.51% and the confidence level is 10.26%, which reveals the relationship between fault and location.

4. Conclusion

China's high-speed railway has accumulated a large number of structured, semi-structured and unstructured production, usage and maintenance data after 8 years operating. For the huge amount of data, it contains valuable security association information of precious EMU parts, whereas the data exhibits the characteristic of high redundancy, wide distribution, and invalidation. This paper aims to establish a fault diagnosis system for high-speed railway. Regarding with the disadvantages in the Apriori algorithm, the Apriori parallel association rules is proposed based on MapReduce function. The analysis system of association rules for fault diagnosis has been established. It is useful for the analysis and prediction of the fault of high-speed railway.

Reference

- Agrawal R, Imieliński T, Swami A (1993). Mining association rules between sets of items in large databases. *Proc. Conf. on Management of Data*. New York: Press ACM.10:207-216
- Park J S, Chen M S, Yu P S (1995) An effective hash-based algorithm for mining association rules. *Proceeding of ACM SIGMOD International Conference on Management of Data*. San Jose, CA, 175-186
- Ashok S, Edward O, Shamkant (1995). An efficient algorithm for mining association rule in large database, *Proceeding of the 21st VLDB conference*, Zurich, Switzerland
- Hannu T (1996) Sampling large databases for association rules, *Proceeding of the 22nd VLDB conference Mumbai* (Bombay), India
- Bi Jianxin, Zhang Qishan (2005) Survey of the algorithms on association rule mining, *Engineering Science*. Vol.7 No.4, Beijing, China
- Wang Zhenyu, Shi Lei Bai, Xiong Fanguan (2002) Research on multiple minimum support degree association rules and mining algorithm. *Mini micro computer system*. 8:971-972
- Hu Hui, Research and Implementation of Mining Association Rules for EMU Failure Data Based on Hadoop[D], Beijing Jiaotong University, Beijing, China, June 2015.

Joint optimization of maintenance preventive “Piggyback” and spare parts policies

T.F. O. Rodriguez. *, L.F.A. Ribeiro. ** and C.A.V. Cavalcante. ***

* Departamento de Eng. de Produção, Universidade Federal de Pernambuco, Brazil: tuliofor@gmail.com

** Departamento de Eng. de Produção, Universidade Federal de Pernambuco, Brazil: lucasfredericoeing@hotmail.com

*** Departamento de Eng. de Produção, Universidade Federal de Pernambuco, Brazil: cristiano@ufpe.br

ABSTRACT: This work addressed the integration of maintenance and spare parts management to establish inventory levels and maintenance interval. It was used the preventive maintenance policy “Piggyback”, which has roots in opportunist policies, and the continuous review system as spare management approach. In this context, it has been modeled a series system of two devices, to study the effects of the cost of preventive maintenance and spare management policy, order, rate and the number of system breakdowns. It was also investigated: (i) factors that make the integration of policies advantageous; and (ii) the consequences that may occur in production procedures.

1. Introduction

The main reasons for a company to maintain an inventory of spare parts are the maintenance actions, in order to restore the system as quickly as possible; so that it performs its intended function (Van Horenbeek et al., 2013). Thus, it is necessary partnership between the areas of maintenance and spare parts management (Panagiotidou, 2013), in order to convey the needs of maintaining the warehouse without delay (Naotokaio et al., 1978).

For this purpose, one must know the type and quantity of parts that must be stored according to the adopted maintenance policy (Gento, 2004). It highlights the importance of having adequate sized inventories, because while excesses generate high costs, low levels can lead to production stoppages, which generate even higher costs.

In this article it is discussed the search for an adequate quantity of parts for stock due to the maintenance policy, optimizing the policies in a synchronous way. This approach was explored by Brown & Proschan (1983); Al-Olayan & Kabir (1996); Sarker (2000); Huiskonen (2001); and Panagiotidou (2013) which demonstrated that, in comparison with sequential optimization, integrated optimization provides better results.

2. Joint “Piggyback” model

Specifically, it was integrated the preventive maintenance piggyback model, proposed by Liang (1985), with the continuous revised model of spare parts management. It is emphasized the flexibility of this approach, which considers the diversity of operating control characteristics of spare parts with the methodology established for maintenance. To verify the effects of the integrated model, it was simulated the effects of variability in the demand for spare parts, and therefore on the total cost per units produced or cost rate, shutdowns rate and emergency or regular requests rate.

2.1. Preventive Maintenance “Piggyback”

The piggyback maintenance policy refers to a particular case of opportunistic maintenance proposed by Liang (1985). This policy applies to a system composed of two units A and B connected in series, such that a failure in equipment A (Piggybackee) follows an exponential distribution. It is also defined a preventive maintenance interval (T) for equipment B (Piggybacker), whose failures follow a Weibull distribution. The “opportunity” is based on a failure of equipment B and T range. There are also model assumptions: the replacement is made with identical equipment whose state is “as good as new”; renewal of the system occurs when the equipment B is replaced; failures on equipment are independent.

Maintenance involves three cases. In the first case, there may be equipment A failures, before failure on equipment B, but all happen before the interval T. In the second case, some of equipment A failures before T and B failure occurs after T but before the subsequent A failure after T, which would offer the opportunity

for preventive replacement of B. Thus, the system is restored in the corrective maintenance of B. The third case features Piggyback policy: the equipment A gives the opportunity to make the preventive maintenance of the equipment B, given that A fails after T without B have been failed.

2.2. Ordering Policies

The spare management model is a set of rules and procedures that let one make some decisions about inventory. In this research, the term refers to spare parts inventory of maintenance that fit the definition of independent demand. For that, it was adopted the methodology of continuous revision (Q system), since it reflects the actual conditions in the area of maintenance with regard to equipment. In this system the inventory item is monitored on every replacement, due to maintenance actions, such that when the stock reaches a certain pre-set level (supply point), a request is emitted. The purchased amount is constant and dependent on resupply point.

2.2.1. Inventory cycle

The letter "Q" represents the number of parts in inventory. The maximum level of inventory (S) - Maximum amount of the item to be stored, depending on the storage space, the inventory keeping costs etc. The minimum level of inventory (s) in turn is dependent on the consumption rate and the regular arrival time request (PR) to ensure that the ordered parts arrive, avoiding the stock shortage. Thus, when $Q = s$ is activated a "PR" with S-s equipment and covering normal delivery times.

In addition, another type of request, called Emergency order (PE), accelerates the entry of items in the inventory in order to avoid a system stop due to lack of inventory, which gives the opportunity of regular order arrives and be stored. This is activated when $Q = 0$, and the quantity to be ordered is pre-established. Fig. 1 is a representation of the situations that can occur in the inventory after each revision. In this example, $S = 7$; $s = 2$; $PR = 5$; $PE = 1$. The production system begins with an operating unit that is not accounted in inventory.

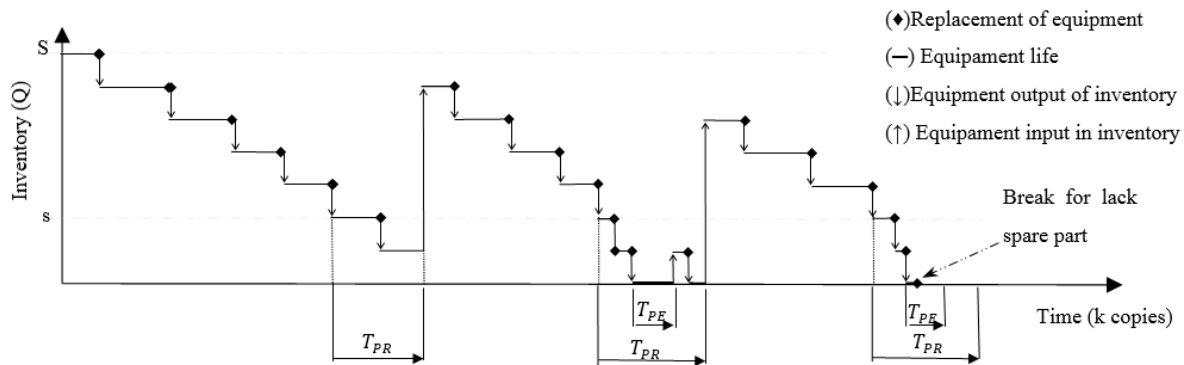


Figure. 1. Profile of B's equipment inventory

2.2.2. Delivery times

Delivery time is the length of time since the order is made up to stocking. The request times are independent of the quantity requested and it is assumed that the waiting time between order and delivery are constant. These times are linked to the type of request, so the regular order time (TPR) is the time that lasts the regular request - it is assumed that the transport used is common (e.g. truck) - and sets the stored minimum amount, which guarantees system availability until the regular application is stored. The time of emergency order (TPE) is less than (TPR) due to the use of special means of transport (e.g. aircraft), which increases the cost of request.

2.2.3. Structure of costs

The following costs are associated with the existence of stock:

The order cost (C_{px}), is the cost of ordering the goods and is the sum of costs incurred from the moment since that the request is made until the time when the parts are stocked. That are staff costs, rent, transportation and costs for revision before the goods are stored multiplied by the number of requests made, dependent on the type of request made and given by the regular order cost ($C_{PR} = C_{pr}N_r$) or the Emergency Order cost ($C_{PE} = C_{ps}N_s$).

Inventory maintenance cost ($C_M = Q \times C_m \times t_e$), is the cost of maintaining equipment in inventory for a certain period. This covers the cost of capital (not available for any activity out of production), cost of storage, which includes the cost of space, insurance, fees, losses, obsolescence or material damage. In the formula, C_m is the cost of maintaining equipment and t_e the length of stay.

Inventory lack cost ($C_p = C_f \times t_p$) reflects the economic consequences of the lack of inventory causing a system break. Some of them are the result of lost sales and / or fines for non-compliance of the product delivery times, and the loss of image and confidence against the consumer. In the formula, C_f is the lack Inventory cost and t_p the time of the break.

The cost of spare management for equipment B is given by ($C_G B = C_{PE} + C_{PR} + C_M + C_p$). For the equipment A, there is the assumption that is provided immediately after a failure; comes down management costs on a single charge called cost per equipment failure ($C_G A = 50$). In turn, the costs entailed by the corrective and preventive maintenance actions are given respectively by $C_{MC} = (tot + td + trpl) \times blr$ and $C_{MP} = trpl \times blr$. Finally, the total cost by integrating Piggyback maintenance model with spare management for N cycles, is $C_{total} = \sum_{i=1}^N (C_G B + C_G A + C_{MC} A + C_{MC} B + C_{MP})$.

3. Model development

The main objective of integration of spare management with the model of preventive maintenance Piggyback is to establish optimal inventory levels (S, s) and preventive maintenance interval (T) from the perspective of the service cost and system availability that consequently reflects the greatest benefit for all.

The simulation process begins by creating random failure times, following the pre-established probability distributions for the equipment A and B, and continues with the definition of preventive maintenance interval (T) and the minimum and maximum levels inventory (S, s). From there, it evaluates the event that could occur in relation to maintenance activities and the implications for the management of spare parts, until it happens the opportunity to renew the system. In the following steps, are calculated: (i) the costs related to maintenance events and inventories instances; (ii) the number of regular and emergency orders; (iii) the number of unscheduled stoppages due to lack of inventory.

This process is repeated for N systems in order to understand the behavior of the decision variables that must act in case of failure, randomly set for each system. Subsequently, these values are displayed as a function of system life and reflect the cost of integrated service rates; regular orders and emergency orders; and system downtime due to lack of inventory.

To achieve the first objective, based on establish the configuration of decision variables (S, s, T) which generate optimum benefit from the viewpoint of the criterion service charge rate. It has been developed an optimization algorithm of the decision variables, in which lies the optimal point related to the service cost for each of the alternatives, as well as the search space in which pre-established limits for the decision variables in order to evaluate different options and thus present the best configuration.

4. Numerical example

Here it is shown the application of the integrated policy compared to sequential to explore the possible advantages and disadvantages of the policies involved. The maintenance data for this numerical example were obtained from Liang (1985) and spare management data were adapted from management situations to spare in photocopy machines. These data are presented in Table 1.

	A		B			B	
	f(t)	Exponencial	Weibull			C_{pr}	60
Maintenance	B	1	3		Inventory	C_{pe}	120
	H	--	67,19	k copies		C_m	0.05
	Θ	30	60	k copies		C_f	1200
	C_F	96.25	99	\$		T_{pr}	54
	C_p	--	11	\$		T_{pe}	18
							k copies

Table 1. Parameters of the models – Simulation

In addition to this, there were used as input parameters of the decision variables the data presented in Table 2 to optimize the cost of service criteria for integrated and sequential models.

S	15	Units	T	0 - 90	k copies
S	0	Units	N	100000	Systems

Table 2. Input data of decisión variables (S, s, T) - Optimization Algorithm

Then, it executes the simulation model using an exhaustive search to find inventory levels and service interval to generate the highest profit for each alternative in reference to the criterion cost of service. The alternatives of optimizing the results are shown in Table 3.

Alternative	Variables (S, s, T)	SC / k copies
Sequential	Main. → Inv. (–, –, 12) → (8, 2, 90)	(4.137) → (2.054)
Joint	Main. + Inv. (13, 3, 17)	(6.273)

Table 3. Results – Optimization

Figure 2 shows the inventory costs for the policies involved. The service cost for the maintenance model for the two alternatives is the same, as the maintenance model depends solely on the range that gives the opportunity to serve the equipment B. After making the optimization, the point which results in lower service cost for maintenance is $T = 12000$ copies with a cost of service rate equal to \$ 4,137 / thousands of copies.

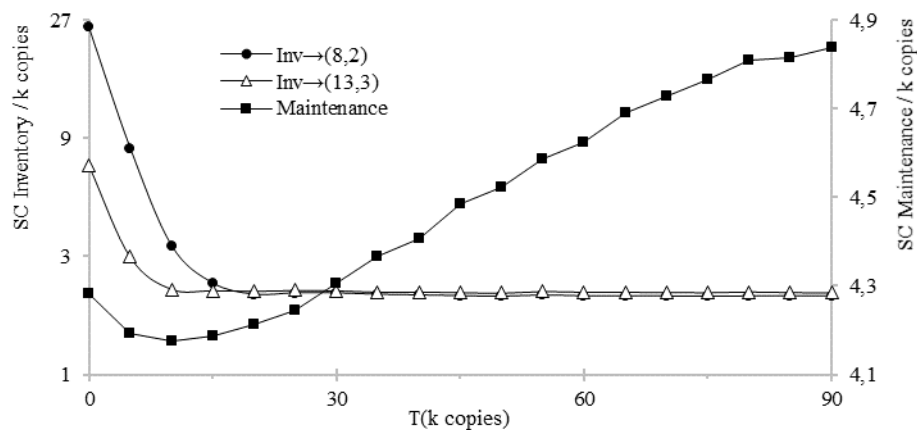


Figure. 2. Service cost rate of inventory and maintenance versus preventive maintenance interval

For the sequential inventory policy model, the optimization defined three decision variables at the point of lowest cost service, resulting in $(S, s, T) \rightarrow (8, 2, 90)$. Therefore, to find the cost of the sequential alternative, the costs involved in maintenance and inventory model are summed, and it is the interval that provides the lowest service cost.

Regarding the integrated alternative it was made the sum of the costs involved, finding three decision variables simultaneously, which represent the best configuration. In Table 4 it is presented the results of sequential policy. In it, there is a single interval of preventive maintenance, which has the lowest cost of service.

Alternative	T (k copies)	S	s	SC / k copies
Sequential	20	8	2	6.375

Table 4. Results – Interval with low service cost rate for sequential policy

Fig. 3 shows the behavior of the integrated and sequential policy for the service costs over the service intervals for the optimal inventory levels (S, s) found for each of the alternatives.

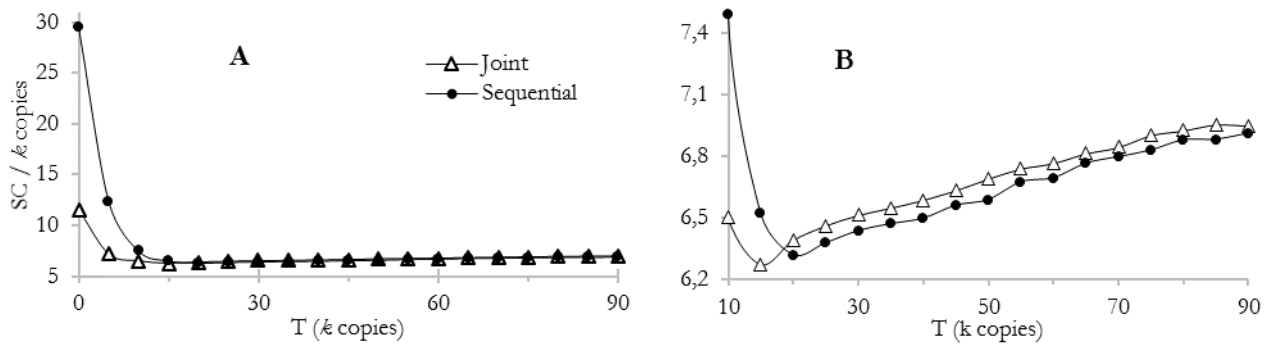


Figure. 3. Service cost rate of sequential and joint policy versus maintenance interval.
(A) complete – (B) Extend

It was found that regardless of the alternative used to decrease maintenance intervals to the optimum point, the cost of service is very high. It is due mainly to the high rate of preventive replacement because they easily met the conditions of Piggyback maintenance policy on case 3.

The different service cost among the evaluated policy is characterized mainly by the different levels of inventory they own. Sequential policy has a higher cost as a result of the low levels found in the individual inventory optimization. Moreover, the model of spare management, particularly treated, reflects, to a single point, the lowest service cost. Thus, low levels of inventory add considerable amount of requests, both regular and emergency and lead to reduced stocking rate, leaving the production system in a vulnerable condition to spare parts shortage greater than the integrated policy. The behavior of requests for treated policies is shown in Fig. 4

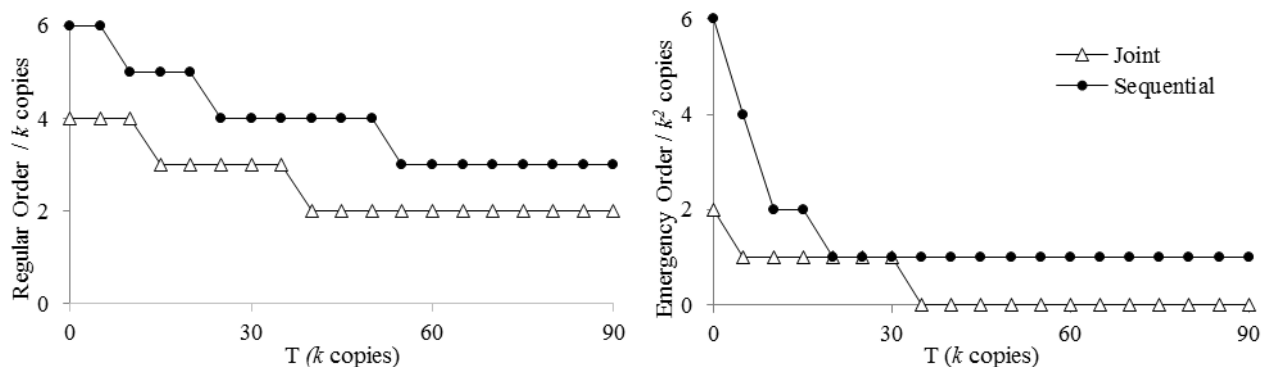


Figure. 4. Regular and Emergency Orders of the sequential and joint policy versus maintenance interval.

The large amount of emergency orders in the first third of the intervals is due to the fact that in the waiting time that a regular request is stored, it should be covered by emergency requests in order to avoid causing a break due to lack of stock. Nevertheless, it presents a large number of systems stopped by the lack of parts in stock, which is not supported by consumption rate (Figure 5).

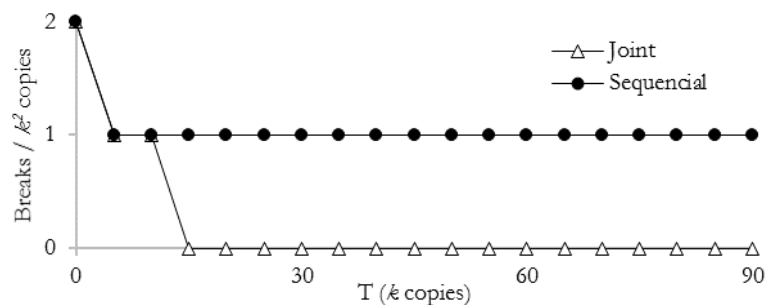


Figure. 5. Breaks due to inventory lack for sequential and joint policy.

In addition, the optimum point of the integrated policy is due mainly to the absence of stops due to lack of inventory and, consequently, the minimum use of emergency requests to meet demand and thus prevent the occurrence of high costs, and to assure system availability.

5. Conclusion

This paper contributes to the development of integrated policies involving maintenance and spares management. Implementation of this policy, beyond establish the balance between maintenance and inventory costs, improves system availability by preventing stops, produced by lack of inventory, and ensures system reliability, by situating it in the midline between unscheduled maintenance and preventive maintenance. In addition, when setting minimum and maximum inventory levels that are subject to the maintenance interval, it ensures that the number of parts in inventory meets the system requirements, minimizing the costs of storage.

6. References

- A. Van Horenbeek, J. Buré, D. Cattrysse, L. Pintelon, and P. Vansteen-Wegen (2013) Joint maintenance and inventory optimization systems: A review. *Int. J. Product. Econ.* 143.2: 446-508.
- Ai-olayan Ahmed S. and Kabir A.B.M. Zohrul (1996) A stocking policy for spare part provisioning under age based preventive replacement. *European Journal of Operational Research* 90 90: 171-181.
- Brown, M. & Proschan, F. (1983) Imperfect repair. *Journal of Applied Probability* 20: 851-859.
- Gento, Angel M. (2004) Decision rules for a maintenance database. *Journal of Quality in Maintenance Engineering* 10: 210 - 220.
- Huiskonen, J. (2001) Maintenance spare parts logistics: Special characteristics and strategic choices. *Int. J. Prod. Econ.* 71.1-3: 125-133.
- Liang, Tom Y. (1985) Optimum Piggyback Preventive Maintenance Policies. *IEEE Transactions on Reliability* R-34.5
- Naoto Kaio; Shunji Osaki. (1978) Optimum Ordering Policies with Lead Time for an Operating Unit in Preventive Maintenance. *IEEE Transactions On Reliability*, 27.4
- Panagiotidou, Sofia. (2013) Joint optimization of spare parts ordering and maintenance policies for multiple identicals items subject to silent failures. *European Journal of Operational Research*
- Sarker, Haque R. (2000) Optimization of maintenance and spare provisioning policy using simulation. *Appl. Math. Model.* 24.10: 751-760.

Buffer Allocation for Continuous Imperfect Production Lines by Simulation and Analytic Network Process

O. M. Topcu* and S.I. Satoglu**

*Toyota Motor Company, Toyota Cad No:2 Arifiye, Sakarya, Turkey. omermeteninmaili@gmail.com

** ITU Isletme Fakultesi, Endustri Muh Bolumu, Macka, Sisli, Istanbul, Turkey. onbaslis@itu.edu.tr

Abstract. The main purpose of this study is the allocation of buffer stock tanks to different products, for the production facilities operating in process type production systems of the fast-moving consumer goods (FMCG) industry. The important aspect of the FMCG sector is that the product variation is high, but the demand for each product type is low. A weekly production plan is constructed which must be satisfied on time with minimum amount of buffer stock. In contrast to the past buffer allocation studies, in process industry, there is a constraint that the liquid products cannot be mixed, unless they are of the same type, within the buffer tanks. Due to the random machine breakdowns, machine repair times, and complex production system structure, simulation is a very useful technique for this problem. A simulation model that measures several criteria reflecting the system performance is designed. After creating this model it is intended to determine the best scenario. Analytic Network Process method has been used as multi criteria decision making method, to evaluate the correlation among the performance criteria, compute the weights of the criteria, and select the best scenario. The proposed methodology has been tested by using a real FMCG factory data.

1. Introduction

Buffer allocation problem aims at finding the optimal buffer sizes to be allocated between the consecutive operations of a manufacturing or assembly line, so that the operations or stations are more independent of one another in time and function (Demir et. al., 2014). It is assumed that the stations of a line have a finite buffer capacity. Jobs cannot leave a station if the downstream station's buffer is full, where the former station is blocked (Rajan and Agrawal, 1994). Besides, parts cannot reach this station from the upstream, if its buffer is full. There is another term called starvation, where a bottleneck station is followed by a faster one, and when the bottleneck does not have enough buffer size, the following station will possibly starve, in other words, it will have parts shortage (Durmusoglu and Satoglu, 2011). Blocking tends to propagate to the upstream where the starvation tends to spread to the downstream operations (Altioek and Melamed, 2010). The resource can become a bottleneck due to failure. Battini et. al. (2009) expressed that buffer size allocation is linked to the reliability of the resources.

Decision of the capacity and number of buffer tanks for the process industries is also a buffer allocation problem. As long as the products are discrete, different types of these can be put into the same buffer space. However, in process industries, there is a constraint that the liquid or continuous products cannot be mixed, unless they are of the same type, within the buffer tanks. Besides, as the product variation increases, the associated problem becomes rather complicated. If the random machine breakdowns, machine repair times, and complex production system structure are also considered, an advanced methodology is required to solve the buffer allocation problem.

The main purpose of this study is development of a novel methodology for solving the buffer allocation problem, and allocation of buffer stock tanks to different products, for the production facilities operating in process type production systems of the fast-moving consumer goods (FMCG) industry. The important aspect of the FMCG sector is that the product variation is high, but the demand for each product type is low. A weekly production plan is constructed which must be satisfied on time with minimum amount of buffer stock, and production plan must be implemented without any backlog.

The simulation technique is employed due to the dynamic and complex nature of the production system. A simulation model that measures several criteria reflecting the system performance is designed. After creating the base-model, several scenarios corresponding to various buffer allocations to different products are generated. It is intended to determine the best scenario. Analytic Network Process method has been used as a multi-criteria decision making method, to evaluate the correlation among the performance criteria,

compute the weights of the criteria, and select the best scenario. The proposed methodology has been tested by using a real FMCG factory data, located in Turkey. Weekly production plans were collected and system constraints about the number and size of the buffer tanks were comprehended. The simulation base-model was constructed where multiple performance criteria are observed for each of the scenario. Based on the results of the simulation runs for all scenarios, Analytic Network Process is applied, where weights of the performance measures are determined, and the scenarios are ranked among each other and the best scenario is selected to represent the best buffer allocation.

The paper is organized as follows: In Section 2 the proposed methodology is explained. Later, the application of the methodology for an FMCG company in Turkey is explained. Finally the Conclusion and future work are presented.

2. The Proposed Methodology

The proposed methodology is based on an Experimental Design that is comprised of alternative buffer allocations to different products and these are represented by the alternative scenarios. While performing the experimental design, the facility's physical constraints must be considered. These are the total number of liquid buffer tanks, the capacity of each tank. Besides, the number of tanks that can be connected to the production stages is also an important constraint. Since the product is continuous some built-in pipes are usually used for the transfer of the work-in process. So, actually all stages cannot be linked to all tanks.

After the experimental design, the simulation model of each scenario should be constructed. In this study, simulation technique is employed, because the production stages have a dynamic behavior, as they break down randomly, their repair times are also random and the demands of the products also vary in time. When the relevant literature is searched, it is observed that the simulation technique has been frequently used under dynamic conditions for bottleneck identification (Durmusoglu and Satoglu, 2011), modeling the maintenance and repair policies (Bal and Satoğlu, 2014), manufacturing system parameter selection such as number of workers, machines, lot sizes (Ertay and Satoğlu, 2011); buffer allocation for the serial unpaced-production lines (Battini et. al., 2009), buffer allocation in serial and parallel production lines (Amiri et. al., 2012), buffer allocation for a small group of unreliable machines (Sørensen, et. al., 2011).

Later, the simulation models of the scenarios must be run, and the performance measures' distributions are estimated. Later, in order to select the most suitable buffer allocation scenario, these must be assessed based on the performance criteria by the using a multi-criteria decision technique called Analytic Network Process (ANP) proposed by Saaty (2005). When the performance criteria are correlated to each other, this technique must be used. Hence, the criteria are weighted and the alternatives are ranked. The alternative scenario that is ranked as the first one is selected as the best buffer allocation for that week's production plan. According to different product demand at each week's production plan, the best buffer allocation result can change. In Figure 1, the stages of the proposed methodology are illustrated.

3. Real Application

The company works in FMCG industry, and the facility of the company produces liquid personal care products. There are three mixers that prepare the mixture. These may break down. Later the mixtures are temporarily stocked in buffer tanks. There are 12 buffer tanks. These are linked to three filling machines where the liquid product is filled into the plastic bottles. In Figure 2, this system structure is illustrated.

According to the recipe of the mixtures and the bottle sizes, products vary. There are more than 60 products that are produced. If the recipes of the products are different from each other, these cannot be put to the same tank.

The time between failures and time to repair distributions of the mixers and filling machines are estimated by the Input Analyzer software. These are not presented here due to space restrictions. Later, 4 scenarios are designed with different number of tanks allocated to supply the filling machines. In table 1, the tanks that are linked to the filling machines are shown for the four scenarios.

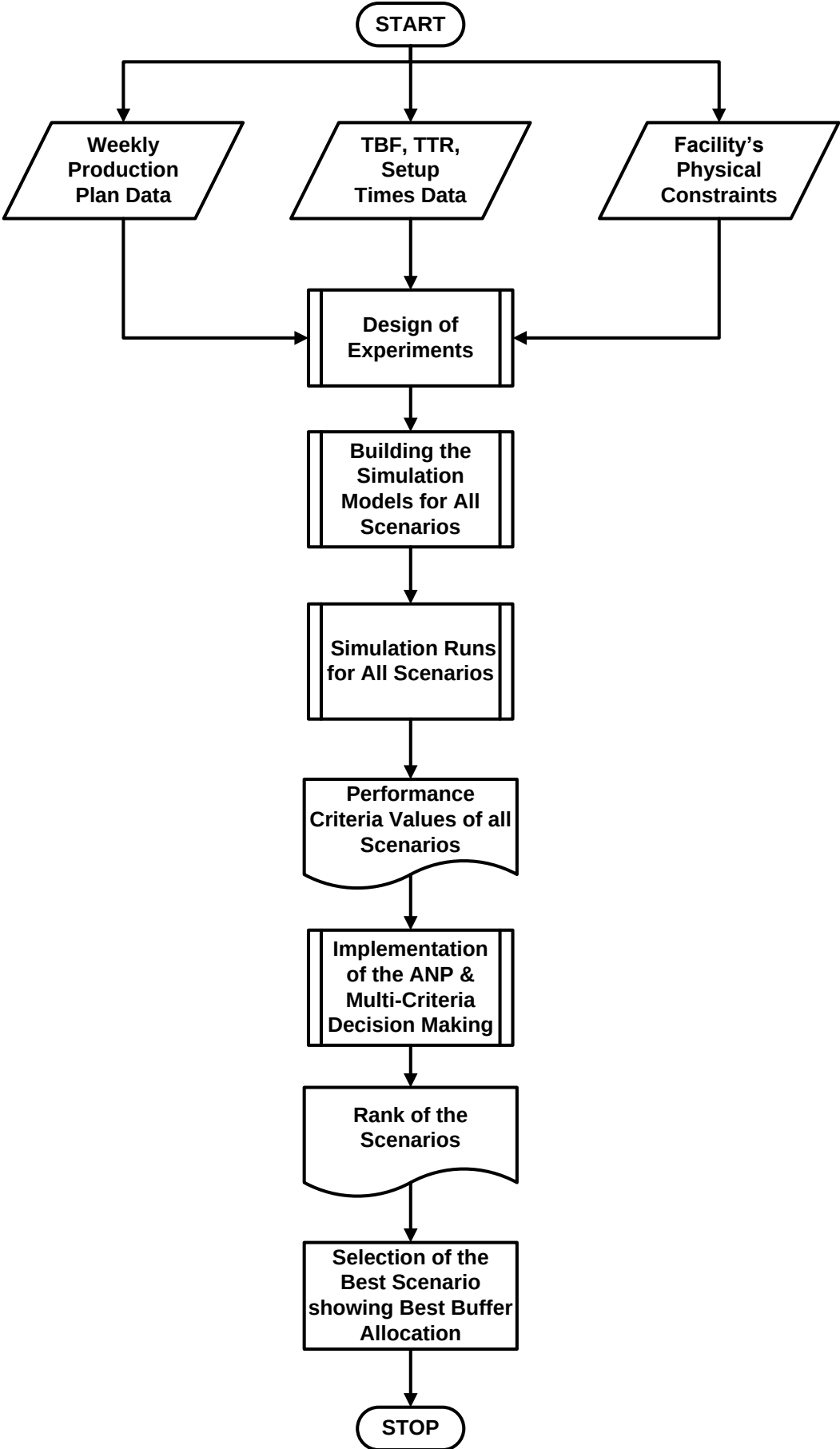


Figure 1. Flowchart of the Proposed Methodology

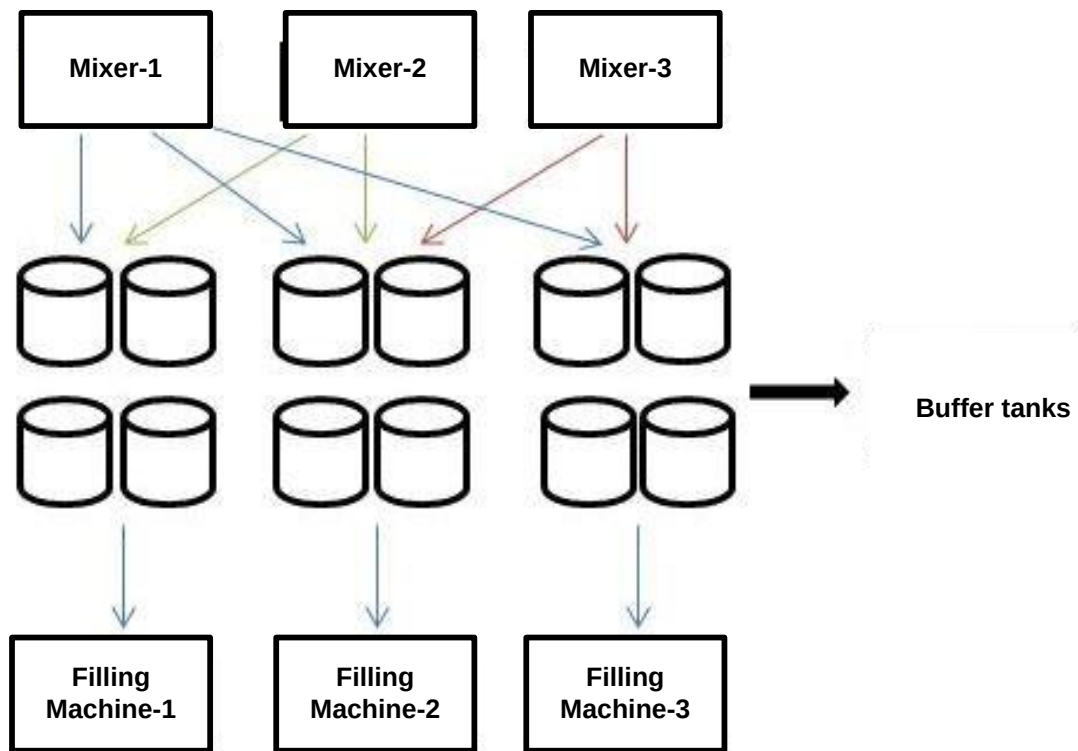


Figure 2. The System Design

Table 1. Experimental Design of the Scenarios

	Number of Tanks Linked to the Filling Machine-1	Number of Tanks Linked to the Filling Machine-2	Number of Tanks Linked to the Filling Machine-3
Scenario-1	4	4	4
Scenario-2	5	3	4
Scenario-3	5	4	3
Scenario-4	6	3	3

The performance criteria to be observed by the simulation models for these scenarios are the time required to complete the total planned production quantity for all products (lead time), average WIP in the tanks (ton/tank), and the state probabilities of the mixers and filling machines. Possible states of these machines are busy, idle, failed and setup. The simulation models were constructed in Arena 10.0® environment. Each scenario has been run for one week period and for five times, to collect sufficient data. The average values of the performance criteria for all scenarios are computed. Table 2 shows all of the scenario results for all performance criteria.

Later, ANP is initiated and the weights of these performance criteria are computed based on the judgements of the experts of the production facility. The ANP methodology is implemented as it is explained in (Saaty, 2005). The weights of the criteria are denoted in Table 3.

According to the ANP methodology, the performance criteria values are normalized, since their measuring units and scales are different from each other. Sum of products of these normalized values of the performance measures with associated weight of the measure are summed for all scenarios and total grade of the scenarios are achieved.

The computations for the total grade of the 1st scenario are presented in Eq. (1).

$$0,173*0,26+0,145*0,25+0,100*0,26+0,072*0,29+0,138*0,24+0,133*0,25+0,099*0,27+0,139*0,26=0,25735 \quad (1)$$

The total grade value for the 2nd Scenario=0,24428

The total grade value for the 3rd Scenario = 0,25129

The total grade value for the 4th Scenario = 0,24747

According to these results, the 1st Scenario is the best one, where four buffer tanks are allocated to each of the three filling machines.

Table 2. Simulation Results of the Scenarios for All Performance Criteria

Scenario #	Mixers Avg. Downtime Ratio (%)	1/Mixers Avg. Downtime Ratio	Normalized Value
1	0,506	1,976284585	0,24
2	0,515	1,941747573	0,24
3	0,492	2,032520325	0,25
4	0,457	2,188183807	0,27
Scenario #	Filling Mach. Avg. Downtime Ratio (%)	1/Filling Mach. Avg. Downtime Ratio	Normalized Value
1	1,161	0,861326443	0,25
2	1,171	0,853970965	0,24
3	1,055	0,947867299	0,27
4	1,185	0,843881857	0,24
Scenario #	Mixers Avg. Setup Time Ratio (%)	1/Mixers Avg. Setup Time Ratio	Normalized Value
1	3,702	0,270124257	0,26
2	3,881	0,25766555	0,25
3	3,632	0,275330396	0,26
4	4,146	0,241196334	0,23
Scenario #	Filling Mach. Avg. Setup Time Ratio (%)	1/Filling Mach. Avg. Setup Time Ratio	Normalized Value
1	8,758	0,11418132	0,29
2	10,833	0,092310533	0,24
3	11,21	0,089206066	0,23
4	10,849	0,092174394	0,24
Scenario #	Mixers Avg. Busy Ratio (%)	1/ Mixers Avg. Busy Ratio	Normalized Value
1	69,835	0,014319399	0,25
2	70,809	0,014122565	0,24
3	65,19	0,015339776	0,26
4	69,481	0,014392355	0,25
Scenario #	Avg. WIP in Tanks	1/Avg. WIP in Tanks	Normalized Value
1	2,51	0,398406375	0,27
2	2,89	0,346020761	0,23
3	2,57	0,389105058	0,26
4	2,75	0,363636364	0,24
Scenario #	Avg. Lead Time	1/Avg. Lead Time	Normalized Value
1	8075	0,000123839	0,26
2	7927,69	0,00012614	0,26
3	8756,45	0,000114202	0,24
4	8258,42	0,000121089	0,25
Scenario #	Mixers Avg. Busy Ratio	1/Mixers Avg. Busy Ratio	Normalized Value
1	72,133	0,013863344	0,26
2	76,541	0,013064836	0,25
3	77,915	0,012834444	0,24
4	75,427	0,013257793	0,25

Table 3. Weights of the Performance Criteria

Performance Criteria	Weight
Mixers Average Busy Ratio	0.173
Filling Mach. Average Busy Ratio	0.145
Mixers Average Setup Time Ratio	0.100
Filling Mach. Average Setup Time Ratio	0.072
Mixers Average Downtime (Fail) Ratio	0.138
Filling Mach. Average Downtime Ratio	0.133
Average WIP in Tanks	0.099
Average Lead Time	0.139

4. Conclusion

In this study, a simulation and ANP-based methodology is proposed for solving the buffer allocation problem especially for the process industries where the products are continuous and there are restrictions in stocking different mixtures of products in buffer tanks. The methodology is applied for the real case of an FMCG facility located in Turkey where personal care products are produced. It is shown that the proposed methodology is capable of finding the most suitable buffer allocation to the filling machines so as to meet the weekly demand.

Although in this study the current buffer tank capacities and number of tanks are considered in real world application, in a future study, scenarios with higher number of tanks that have smaller capacities can be investigated, since the high product variety complicates the buffer allocation problem in this industry.

References

- Altiock T, Melamed B (2010) *Simulation modeling and analysis with Arena*. Academic Press.
- Amiri, M., Mohtashami, A. (2012). Buffer allocation in unreliable production lines based on design of experiments, simulation, and genetic algorithm. *International Journal Advanced Manufacturing Technology*, **62**, 371 – 383
- Bal A, Satoglu, SI (2014) Maintenance Management of Production Systems with Sensors and RFID: A Case Study. In *Global Conference on Engineering and Technology Management (GCETM)* (Calisir Eds.), pp.82-89.
- Battini D, Persona A, Regattieri A (2009) Buffer size design linked to reliability performance: A simulative study. *Operational & Industrial Engineering*, **56**, 1633 – 1641.
- Demir L, Tunali S, Eliyi DT (2014) The state of the art on buffer allocation problem: a comprehensive survey. *Journal of Intelligent Manufacturing* **25**, 371-392.
- Durmusoglu MB, Satoglu SI (2011) Axiomatic design of hybrid manufacturing systems in erratic demand conditions. *International Journal of Production Research* **49**, 5231-5261.
- Ertay T, Satoğlu SI (2012) System parameter selection with information axiom for the new product introduction to the hybrid manufacturing systems under dual-resource constraint. *International Journal of Production Research*, **50**, 1825-1839.
- Rajan R, Agrawal R (1994) Cyclic networks with general blocking and starvation. *Queueing Systems* **15**, 99-136.
- Saaty TL (2005) *Theory and Applications of the Analytic Network Process*, RSW Publications, Pittsburg, USA
- Sörensen, K., Janssens G. K. (2011). Simulation results on buffer allocation in a continuous flow transfer line with three unreliable machines. *Advances in Production Engineering & Management*, 6(1), 15–26.

Redundancy Allocation for Series-Parallel Warm-Standby Systems with Interval-Valued Lifetime Deceleration Factor

Wei Wang^{*, **}, Junlin Xiong^{*} and Min Xie^{**}

^{*}Department of Automation, University of Science and Technology of China, China. wwwayne@mail.ustc.edu.cn, xiong77@ustc.edu.cn.

^{**}Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong. minxie@cityu.edu.hk.

Abstract. This paper investigates the redundancy allocation problem (RAP) for a series-parallel warm-standby system with uncertain-but-bounded lifetime deceleration factors. The factors are studied by using interval analysis and represented as interval numbers with upper bounds and lower bounds. The system reliability is obtained and formulated as interval value by using interval arithmetic. Then, this paper presents a new definition of order relation for interval numbers according to decision makers' preference. Based on the order relation, we can select the best alternatives for the RAP. We develop a genetic algorithm (GA) for solving the RAP of series-parallel warm-standby systems. Finally, an illustrative example is presented.

1. Introduction

Redundancy allocation is a widely used technique for improving the reliability of a system. According to the failure characteristics of components in the standby mode, there are three standby types: hot-standby, cold-standby and warm-standby. Warm-standby is a fault-tolerant design technique developed as a compromise between hot-standby and cold-standby approaches in terms of power consumption and recovery time. Both hot- and cold-standby are essentially special cases of the warm-standby model. Compared with hot-standby spares, the warm-standby spares do not consume much power when they are in standby mode, also they suffers milder working environment and reduced failure rate. While compared with cold-standby spares, the warm-standby spares do not need long recovery time when they are activated into working mode.

Many studies have been carried out on the reliability evaluation of warm-standby system with different structures. She and Pecht (1992) presented a calculation method and a close-form expression for general k -out-of- n warm-standby systems. Amari et al. (2012) studied the reliability characteristics of a k -out-of- n warm-standby system with identical components subject to exponential lifetime distributions. Moreover, reliability optimization problems of warm-standby systems also achieve a lot of attention. Tannous et al. (2011) proposed and compared two approaches for solving RAP of series-parallel warm-standby systems. Yun and Cha (2010) studied an optimal problem for a two-unit system with one active component and one cold-standby component, and the cold-standby component will be switched into warm-standby mode. Levitin et al. (2014; 2015) presented a series of studies on optimization problems of warm-standby systems.

In the existing literature, warm-standby components in the standby mode are usually studied by using the ideas about accelerated life tests. Under different stress conditions, there are different methods that can be used to model the failure rates of warm-standby components, such as the proportional hazards model (Ghasemi et al., 2010), the accelerated failure time model (Blischke and Murthy, 2011), and the virtual age model (Finkelstein, 2009; Zhai et al., 2015). In the first two models, the failure rates of warm-standby components are reduced through some factors which are known as exact number between 0 and 1. In the virtual age model, a non-decreasing function is presumed to given to model the components behaviors under standby mode. However, due to complicated environment, system complexity, insufficient data, or limited resource, it is a difficult task to determine the exact values of the reduced failure rates, the lifetime deceleration factors, or the parameters in these non-decreasing functions. When these values can only be estimated as uncertain-but-bounded numbers, there is not much work toward analyzing the systems reliability modelling and optimization with warm-standby redundancies.

This paper investigates the RAP for a series-parallel warm-standby system with uncertain-but-bounded lifetime deceleration factors. The factors are represented as interval numbers with upper bounds and lower bounds. Interval analysis theory (Moore et al., 2009) is used to deal with the interval numbers. The system reliability is also formulated as interval-valued by incorporating interval arithmetic with the existing methods. Then, this paper proposes a new definition of order relation for interval numbers according to decision makers' preference. By using the order relation, we can select the better alternatives for the RAP. In this paper, we propose a GA based approach for solving the RAP for series-parallel warm-standby systems.

The method proposed in this paper provides system designers a possibility that they can make decisions even though the considered warm-standby systems have interval-valued deceleration factors. This method improves the robustness of reliability designing for the warm-standby systems.

2. Model Description

In this section, we firstly present a brief introduction to interval arithmetic and give the definition of an order relation based on decision makers' preference is defined for comparing interval numbers. Then we discuss the characteristics of warm-standby components under milder stress with interval-valued lifetime deceleration factors. Thirdly, we derive the reliability function of the series-parallel warm-standby system. Finally, the RAP for series-parallel warm-standby systems is formulated.

2.1 Interval Arithmetic

In interval arithmetic (Moore et al., 2009), an uncertain variable a is represented as an interval number $[a] = [\underline{a}, \bar{a}]$ with $\underline{a} \leq a \leq \bar{a}$, where $\underline{a}$ and $\bar{a}$ are the lower bound and the upper bound of a , respectively. If $\underline{a} = \bar{a}$, then a is a real number. The basic arithmetical operations of two interval variables $[a] = [\underline{a}, \bar{a}]$ and $[b] = [\underline{b}, \bar{b}]$ ($\underline{a} \geq 0, \underline{b} \geq 0$) are defined as

$$[a] \otimes [b] = \{x \otimes y | \underline{a} \leq x \leq \bar{a}, \underline{b} \leq y \leq \bar{b}\}. \quad (1)$$

where $\otimes \in \{+, -, \times, \div\}$ and in the case of division that $\bar{b} \neq 0$ and $\underline{b} \neq 0$. For more operations and details of interval arithmetic, one can refer to Moore et al. (2009).

Since there are three types of relations for any two intervals: non-overlapping, partial overlapping, and inclusion, the order relations between real numbers are not applicable. In order to compare any two interval numbers of all these three types, we propose a new definition of order relation between intervals with considering the measurement of decision maker's preference (denoted as ρ_0).

Definition 1. The order relation $\geq_{max}$ between intervals $[a] = [\underline{a}, \bar{a}]$ and $[b] = [\underline{b}, \bar{b}]$ for maximization problems is defined as:

For the non-overlapping type ($\bar{a} < \underline{b}$) and the partial overlapping type ($\underline{a} \leq \underline{b} \leq \bar{a} \leq \bar{b}$),
 $[a] \geq_{max} [b]$ if $\underline{a} \geq \underline{b}$ and $\bar{a} \geq \bar{b}$;

For the inclusion type ($\underline{b} < \underline{a} \leq \bar{a} < \bar{b}$),
 $[a] \geq_{max} [b]$ if $\rho \geq \rho_0$, where $\rho = (\underline{a} - \underline{b}) / (\bar{b} - \bar{a})$.

2.2 Warm-Standby Components with Interval-Valued Lifetime Deceleration Factor

Let $F(t)$ and $F_w(t)$ denote the distributions of the lifetimes of a component under the operational mode and the warm standby mode, respectively. When the component is in the warm-standby mode, it suffers milder working stress than the component in operational mode. Therefore, for $\forall t \geq 0$,

$$F_w(t) \leq F(t). \quad (2)$$

Based on the concept of virtual age, a general model has been widely used and shown as

$$F_w(t) = F(\gamma(t)), \quad (3)$$

where $\gamma(t) \leq t$ is a non-decreasing function, $t \geq 0$ and $\gamma(0) = 0$. For a special case of linear accelerated failure time model, $\gamma(t) = \alpha t$, $0 < \alpha < 1$. Suppose there is an exponential component with constant failure rate λ , the stress in warm-standby mode has the effect of expanding time through $\gamma(t) = \alpha t$. So, if the component in warm-standby mode, it will suffer a failure rate of $\alpha\lambda$.

Considering the difficulties of the complicated and dynamic environment, system complexity, insufficient data, or limited resource, it is overly difficult to determine the exact values of parameters in the function $\gamma(t)$. When given the interval-valued factors, the function $\gamma(t)$ can be represented as

$$[\gamma(t)] = \gamma(t; [\alpha_1], [\alpha_2], \dots), \quad (4)$$

where $[\alpha_1], [\alpha_2], \dots$ are the interval-valued parameters of $\gamma(t)$. For the linear case, $[\gamma(t)] = [\alpha]t$.

2.3 Reliability Function of Series-Parallel Warm-Standby System

For a series-parallel warm-standby system, there are S parallel subsystems connected in series. Subsystem i is a 1-out-of- n_i warm-standby system. In each subsystem, all components are assumed to have exponential distribution with identical failure rate. The system fails only if one of the subsystems fails. A subsystem fails when all the components in this subsystem fail. Hence, the reliability of the whole series-parallel system can be obtained as the product of the reliabilities of all the S subsystems,

$$R_{sys}(t) = \prod_{i=1}^S R_i(t). \quad (5)$$

Following She and Pecht (1992) and considering the interval-valued lifetime deceleration factor $[\alpha_i]$, the reliability function of subsystem i , which is also a 1-out-of- n_i warm-standby system, can be formulated by using the basic interval arithmetic,

$$[R_i(t)] = \frac{1}{([\alpha_i]\lambda_i)^{n_i-1}} \times \sum_{j=0}^{n_i-1} (-1)^j \times \left\{ \prod_{k=0, k \neq j}^{n_i-1} (1 + k[\alpha_i])\lambda_i \right\} \times \frac{\exp(-(1+j[\alpha_i])\lambda_i t)}{(n_i-1-j)! \times j!}. \quad (6)$$

where n_i is the number of components in subsystem i ; λ_i is the full failure rate of component in working mode; we assume the system has perfect fault detection and switching functions.

Since the system consists of S parallel subsystems connected in series, the reliability of the whole series-parallel system is obtained as,

$$[R_{sys}(t)] = \prod_{i=1}^S [R_i(t)]. \quad (7)$$

Remark: Though expression (6) provides a theoretically expression for reliability of subsystem i , however, due to the numerical round-off errors, the expression may leads to unstable computational results for large values of n_i and for small values of $[\alpha_i]$. Therefore, in order to avoid the unstable computational results, one can use the algorithm of reliability evaluation for a k -out-of- n warm-standby system suggested by Amari et al. (2012).

2.4 RAP Formulation

The series-parallel system consists of S parallel subsystems connected in series. The system designer can use different types of components in each subsystem. And different types of components can have different failure rates and different consume of resource. The purpose of RAP is to allocate redundancies to maximize the system reliability while meet the resource limitations or cost constraints. Hence, the RAP is formulated as follows,

$$\begin{aligned} \text{Max} \quad & [R_{sys}(t; \mathbf{n})] = \prod_{i=1}^S [R_i(t; n_i)] \\ \text{s.t.} \quad & g_k(t; \mathbf{n}) \leq b_k, \forall k \in I, \end{aligned} \quad (8)$$

where n_i is the number of components in subsystem i which is also the decision variable in this problem; $\mathbf{n} = (n_1, n_2, \dots, n_S)$ is the solution vector; $g_k(\cdot)$ is a function for consumed resource; b_k is the maximum value of the k th resource constraints; and I is the set of all possible resource constraints. The order relation described in Definition 1 for maximization problems is used in problem (8) for comparing interval values.

3. RAP Solving Using Genetic Algorithm

This section develops a GA to solve the RAP of a series-parallel warm-standby system with interval-valued lifetime deceleration factors. A brief description of the algorithm is given in this section. Here we define the solution encoding and the specific GA operators used in this algorithm. Each solution is implemented as a vector of the number of components in each subsystem and given by,

$$\mathbf{n} = \{n_1, n_2, n_3, \dots, n_S\}. \quad (9)$$

The initial generation is created by randomly selecting a set of solution vectors. The size of the initial population should be reasonable. It should not be small to have a genetic diversity; meanwhile, it should not be too large, because increasing the population size will increase the computational time.

Fitness Function: The fitness function of each individual in the population is obtained by using penalty function technique. It is defined as the sum of the objective function and a penalty value,

$$[fitness(t; \mathbf{n})] = [R_{sys}(t; \mathbf{n})] + \delta \cdot [-M, -M + 1], \quad (10)$$

where M is a sufficiently large positive number; for unfeasible solutions, $\delta = 1$; for feasible solutions, $\delta = 0$. The values of fitness functions are also interval-valued.

Selection Operator: There are several existing methods for selection in GA. A commonly used process is the roulette selection. This operator is used to attain the parents in our algorithm.

Crossover Operator: Crossover operator is applied to exchange genetic characteristics of the parent. A $1 \times S$ binary vector (crossover mask) is randomly generated in every time the crossover operator performs. All 1s in the crossover mask mean the genes in the same positions of the parent's chromosomes should be exchanged, and all 0s mean no action.

Mutation Operator: Mutation operator performs random perturbations to selected chromosomes, in order to jump out of local optimal solution. In this operator, a dual mutation technique with global mutation and local mutation is used. For the global mutation, a $1 \times S$ vector is generated randomly with the value of its elements in (0, 1), the gene of the selected chromosome in the position, where the value is lower than the global mutation rate, should be altered randomly. For the local mutation, once a solution is selected among the current best solutions at the local mutation rate, a new solution is generated locally by using a random walk. Random walk means the gene of the selected chromosome are altered in a small range.

Termination Criteria: Above operators are repeated until the maximum number of generations is reached.

4 Illustrative Examples

This section presents an example to illustrate the implementation of the proposed GA for solving the RAP of a series-parallel warm-standby system with interval-valued lifetime deceleration factors. The system consists of 14 subsystems connected in series. The maximum number of components in each subsystem is 8. The mathematical model of the RAP of the warm-standby system applying the order relation described in Definition 1 is represented as follows:

$$\begin{aligned} \text{Max} \quad & [R_{sys}(t; \mathbf{n})] = \prod_{i=1}^{14} [R_i(t; n_i)] \\ \text{s.t.} \quad & \sum_{i=1}^{14} c_i n_i \leq C, \\ & 1 \leq n_i \leq 8: \text{ integer, } i = 1, 2, \dots, 14. \end{aligned} \quad (11)$$

The characteristics of the components are adapted from Tannous et al. (2011) and shown as follows,

The full failure rate λ_i of the components in each of the 14 subsystems is given by:

$$\lambda_i = \{2, 4, 4, 6, 2, 3, 4, 6, 2, 1, 5, 4, 2, 8\} \times 10^{-2} \text{ failure/hour};$$

The deceleration factor $[\alpha_i]$ of the components in each of the 14 subsystems is given by:

$$\begin{aligned} [\alpha_i] = \{ & [0.1272, 0.1477], [0.1309, 0.1407], [0.1308, 0.1422], [0.1290, 0.1477], [0.1260, 0.1471], \\ & [0.1384, 0.1498], [0.1349, 0.1520], [0.1334, 0.1484], [0.1319, 0.1465], [0.1297, 0.1519], \\ & [0.1337, 0.1447], [0.1290, 0.1512], [0.1313, 0.1418], [0.1257, 0.1451]\}; \end{aligned}$$

The cost c_i of the components in each of the 14 subsystems is given by:

$$c_i = \{20, 25, 30, 35, 15, 45, 20, 40, 60, 10, 30, 75, 55, 45\}.$$

In the computation of this example, the GA is performed with a population of 50 individuals, randomly but diversely initialized. The crossover rate is set to 0.95. The global mutation rate is set to 0.3 and the local mutation rate is 0.45. A maximum number of generations will be set as the termination criteria of the GA.

Table 1 shows the best configuration of components in each subsystem with different maximum system cost: $C = 2000, 2500$ and 3000 and different maximum number of generations: $max_gen = 20, 50$ and 100 . The mission time is 10 hours and $\rho_0 = 1$. The computational time is also presented. We can see from Table 1 that the higher maximum system cost lead to the higher system reliability. Increasing the maximum number of generations will significantly improve the searching results but on the other hand will increase the computational time. Figure 1 shows the convergence process of the best fitness value in each generation for $t = 10$ hours, $C = 2000$ and $\rho_0 = 1$. The optimal system reliability is achieved after about 60 generations. The curves show that the proposed GA consistently converged to the optimal solution.

Table 2 shows the best obtained solutions of the RAP with different mission time varying from 5 hours to 40 hours. The maximum system cost is set as 2000 and the maximum number of generations is set to 100.

The preference level of decision makers is also set to $\rho_0 = 1$. It can be observed from Table 2 that the system reliability decreases gradually as mission time goes on.

Table 1. Best obtained solutions for different C and max_gen with $\rho_0 = 1$.

C	max_gen	Best Solution	Fitness/Reliability	Cost	Computational Time
2000	20	4,4,4,4,4,3,5,5,3,3,4,4,3,6	[0.974588, 0.975930]	2000	51.8544
	50	4,4,4,5,4,3,4,5,3,3,5,4,3,5	[0.978275, 0.979565]	2000	115.1099
	100	4,4,4,5,3,4,4,5,3,3,4,4,3,5	[0.978488, 0.979854]	2000	226.2165
2500	20	4,5,5,6,5,5,5,6,4,4,5,5,4,6	[0.997103, 0.997415]	2500	50.7811
	50	4,5,5,6,4,4,5,6,4,5,5,5,4,7	[0.997176, 0.997424]	2495	119.5439
	100	4,5,5,6,5,4,5,6,4,4,5,5,4,7	[0.997278, 0.997517]	2500	241.6060
3000	20	5,6,6,8,5,5,7,7,4,6,6,6,5,8	[0.999611, 0.999653]	2995	53.0194
	50	6,6,6,7,5,5,6,7,5,4,6,6,5,8	[0.999675, 0.999717]	3000	126.7630
	100	5,6,6,7,5,5,7,7,5,4,6,6,5,8	[0.999688, 0.999728]	3000	250.1008

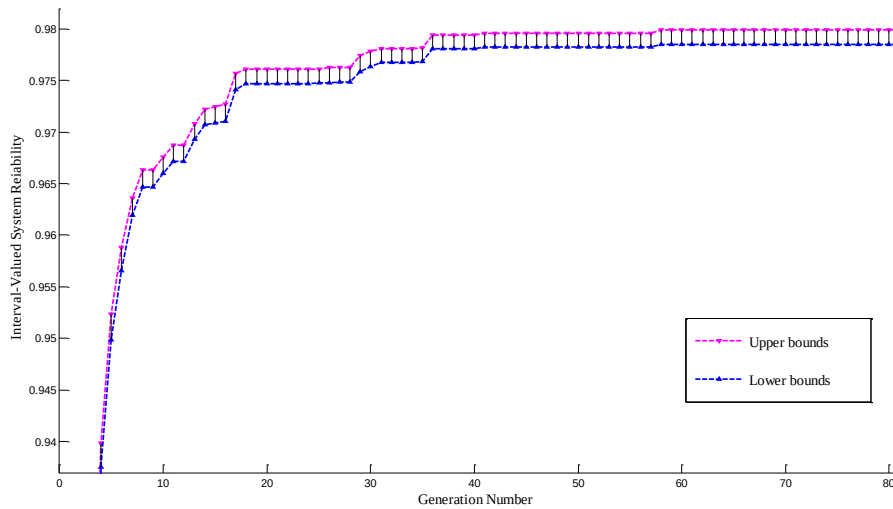


Figure 1. Convergence process of the best fitness in each generation.

Table 2. Best obtained solutions for different mission time with $\rho_0 = 1$ and $C = 2000$.

Mission Time	Best Solution	Fitness/Reliability	Cost	Computational Time
5	4,4,4,5,3,4,4,5,3,3,4,4,3,5	[0.998220, 0.998323]	2000	223.5378
10	4,4,4,5,3,4,4,5,3,3,4,4,3,5	[0.978488, 0.979854]	2000	226.2165
15	3,4,4,5,4,3,5,5,3,3,5,4,3,5	[0.915052, 0.920091]	2000	226.7203
20	3,4,4,5,3,4,5,5,3,3,5,3,3,6	[0.793933, 0.804598]	2000	224.7988
25	3,4,4,5,3,4,5,5,3,3,5,3,3,6	[0.629811, 0.646849]	2000	226.6991
30	3,4,5,5,4,3,5,5,3,3,5,3,3,6	[0.449203, 0.469596]	2000	224.9335
35	3,4,5,5,4,3,5,5,3,3,5,3,3,6	[0.286224, 0.306315]	2000	231.3515
40	3,5,5,6,4,3,5,5,2,3,5,3,3,6	[0.164320, 0.179934]	2000	225.6127

Table 3. Three different solutions for decision makers with different preferences.

Solution No.	Solution	Fitness/Reliability	Cost	ρ (vs. Solution 3)
1	3,4,4,4,3,3,4,5,3,5,5,4,3,6	[0.97326091, 0.97469603]	1995	11.47
2	3,4,5,5,4,4,4,5,3,3,4,3,3,6	[0.97318756, 0.97458656]	1995	2.36
3	5,4,4,5,3,4,5,5,3,3,5,3,3,5	[0.97315958, 0.97470486]	1995	

Table 3 presents three different configurations at mission time of 10 hours and maximum cost of 2000. Based on the order relation in Definition 1, it is obvious that Solution 1 is better than Solution 2, because both the lower bound and the upper bound of the system reliability in Solution 1 are superior to that in

Solution 2. For Solution 1 vs. Solution 3, $\rho = 11.47$, that means only the very most optimistic decision makers with preference $\rho_0 > 11.47$ will select Solution 3 as the better alternative. For Solution 2 vs. Solution 3, $\rho = 2.36$, that means, more optimistic decision makers would like to select Solution 3 rather than Solution 2, even though they will take risk for a possible minimum reliability. The example illustrates that with the increase of ρ_0 the decision makers become more optimistic and adventurous.

5 Conclusions

This paper considers the RAP of series-parallel systems with warm-standby components having interval-valued lifetime deceleration factors. Interval analysis is used to deal with the interval-valued lifetime deceleration factors. The problem is formulated by considering the components having exponential time-to-failure distribution. A new order relation according to decision makers' preference is defined for comparing interval values. A GA based approach has been implemented to obtain the optimal design configurations. The results show that the interval analysis is a useful tool to deal with warm-standby system optimization problems with interval-valued lifetime deceleration factors.

The future work includes extending the presented methodologies for solving RAP of series-parallel warm-standby systems having any arbitrary failure distributions, and incorporating interval analysis into other standby optimization problems, such as k -out-of- n cold standby, mixed cold-hot standby, etc.

References

- Amari SV, Pham H, Misra RB (2012) Reliability characteristics of k -out-of- n warm standby systems. *IEEE Transactions on Reliability* **61**(4), 1007-1018.
- Bhattacharyya GK, Soejoeti Z (1989) A tampered failure rate model for step-stress accelerated life test. *Communications in Statistics-Theory and Methods* **18**(5), 1627-1643.
- Blischke WR, Murthy DP (2011) *Reliability: Modeling, Prediction, and Optimization*. John Wiley & Sons.
- Finkelstein M (2009) Virtual age of non-repairable objects. *Reliability Engineering & System Safety* **94**(2), 666-669.
- Ghasemi A, Yacout S, Ouali MS (2010) Evaluating the reliability function and the mean residual life for equipment with unobservable states. *IEEE Transactions on Reliability* **59**(1), 45-54.
- Levitin G, Xing L, Dai Y (2014). Mission cost and reliability of 1-out-of- n warm standby systems with imperfect switching mechanisms. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **44**(9), 1262-1271.
- Levitin G, Xing L, Dai Y (2015). Non-homogeneous 1-out-of- N warm standby systems with random replacement times. *IEEE Transactions on Reliability* **64**(2), 819-828.
- Moore RE, Kearfott RB, Cloud MJ (2009) *Introduction to Interval Analysis*, SIAM.
- Peng Y, Dong M, Zuo MJ (2010) Current status of machine prognostics in condition-based maintenance: a review. *The International Journal of Advanced Manufacturing Technology*, **50**(1-4), 297-313.
- She J, Pecht MG (1992) Reliability of a k -out-of- n warm-standby system. *IEEE Transactions on Reliability* **41**(1), 72-75.
- Tannous O, Xing L, Rui P, Xie M, Ng SH (2011) Redundancy allocation for series-parallel warm-standby systems. In *Proc. IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 2011. IEEE. pp.1261-1265.
- Yun WY, Cha JH (2010) Optimal design of a general warm standby system. *Reliability Engineering & System Safety* **95**(8), 880-886.
- Zhai Q, Yang J, Peng R, Zhao Y (2015) A Study of optimal component order in a general 1-out-of- n warm standby system. *IEEE Transactions on Reliability* **64**(1), 349-358.

Predicting Discounted Warranty Cost Over Time Considering Failed-But-Not-Reported Events

Xin WANG*, Zhisheng YE*, Wei XIE**, Loon Ching TANG*

**Department of Industrial & Systems Engineering, National University of Singapore*

email: xin.wang@u.nus.edu, iseyez@nus.edu.sg, isetlc@nus.edu.sg

*** Department of Industrial Engineering, South China University of Technology*

email: bmwxie@scut.edu.cn

Abstract Warranty claims incur substantial costs to manufacturers. The analysis of warranty cost is of great interest in reserve preparation. It is not accurate or economically effective to reserve money for future claims based on the analysis of life cycle warranty cost. Being able to predict warranty cost in a flexible period of time could lead to reduced liquidity risks and improved warranty service efficiency. If a product fails in the field during the warranty period, a warranty claim will be incurred only when its user reports the failure to the manufacturer. This paper focuses on the evaluation of discounted warranty cost over different periods of time, with the consideration of customer report behavior. We start with mathematical formulations of various factors, following which we discuss the modeling process of the discounted cost. Expectation and variance of the discounted cost are derived under a pro-rata policy. The impacts of warranty length and customer report behavior are explored. The statistical model for estimation is briefly discussed. The analysis of the discounted warranty cost over a flexible period is useful in planning future warranty services and budgets over a specific time period.

1. Introduction

Most products carry a warranty that protects consumers from problems with the goods they buy. Administering and serving warranty claims brings substantial costs. Typically, warranty costs account for 2% to 15% of product net sales (Blischke et al. 2011). To ensure quality of warranty service and reduce liquidity risk, manufacturers often need to set aside an appropriate amount of money for future warranty obligations. The cost of depleting a warranty reserve is great as the manufacturers have to look for emergent funding at a high interest rate. On the other hand, excess warranty reserve is not desired as well because it incurs opportunity cost (Gurgur 2011). Generally, the warranty reserve is set based on an estimate of potential warranty expense. Reasonable estimation of warranty cost is, therefore, of great importance to manufacturers.

Most of research on warranty cost modeling emphasize more on product and manufacturer related factors such as design reliability, system structure (Bai and Pham 2004), and warranty policy Polatoglu and Sahin (1998). One commonly overlooked factor is customer's report behavior. An implicit assumption in most of previous studies is that a failure will definitely result in a warranty claim. However, there is no guarantee that consumer will report the failure and return the failed unit to the same producer. In reality, the failed-but-not-reported (FBNR) phenomenon is quite common especially for products that are not expensive. Customers may not like the original product or do not bother wasting money or time for a repair or a replacement. Such FBNR phenomenon could lead to a great reduction in warranty claims and costs if the FBNR proportion is high. Then, the over-reserved money will be a waste in this case and cause opportunity cost. Patankar and Mitra (1995) proposed two time-dependent execution functions to represent customer report behavior and examined effects of FBNR phenomenon on warranty cost per unit. Wu (2011) further derived warranty cost with the consideration of both FBNR behavior and non-failed-but-not-reported behavior.

Numerous studies have concentrated on calculating warranty cost per unit and obtained total cost for a single lot sale by multiplying the cost per unit and the size of the lot sale (Menke 1969; Amato and Anderson 1976). Such method of evaluating warranty cost over a certain period of time is only applicable when all units are sold at the same time. As units are usually put on sale at different times, dynamic sales processes should be taken into consideration in warranty cost estimation. Ja et al. (2002) determined mean and variance of total discounted warranty cost for the stochastic sales over the life cycle of the product. This life cycle cost analysis helps manufacturers to evaluate a warranty program and plan for the amount of warranty reserve at the

beginning of the sales period. However, it is a great waste to reserve a large quantity of money for the whole life cycle at the beginning of the sales period. Another concern is regarding the accuracy of such a long-term forecasting. A more reasonable and flexible way is to predict warranty expense progressively and to reserve money for a given period of time. Xie and Ye (2016) derived expectation and variance of discounted aggregate cost for any period starting from the time that the product is introduced to the marketplace. For a future period not starting from product launch time, even though the expectation of cost can be readily obtained, the variance is not available from their model. Their model cannot account for customer report behavior as well.

Our objective is to calculate both mean and variance of the total discounted warranty cost during any period of time. Especially, the FBNR phenomenon, an important factor that directly influences the number of warranty claims, are considered in our model. We propose a continuous function to characterize the non-decreasing FBNR behavior. Other factors including stochastic sales process, warranty policy, and methods of discounting are also discussed in the modeling of the discounted warranty cost. Specifically, a nonhomogeneous poisson process is used to describe the sales process. Analysis of warranty cost is carried out under a pro-rata warranty policy and a continuous discounting method. Besides providing general expressions for the mean and variance of the discounting cost, we also derived explicit forms for some special cases. The knowledge of the expected discounted warranty cost is useful for manufacturers to determine the size of a warranty reserve. However, the expected value alone cannot adequately inform the likely size of the total cost. As a good measurement of uncertainty, the variance of the warranty cost could help managers to determine the risk of having insufficient or excess warranty reserve. Our approach is useful for manufacturers to plan for warranty reserve over any finite time horizon.

2. Model Formulation

Let $TDC(t_s, t_e)$ denote the total discounted warranty cost over a given period of time $(t_s, t_e]$ where $t_s \leq t_e$. This means $TDC(t_s, t_e)$ is obtained by first discounting expense incurred within $(t_s, t_e]$ back to the starting time t_s and then adding up all the discounted values. It should be noted that the warranty expense arises for items that are still under warranty and will fail and be reported within the study period. Essentially, $TDC(t_s, t_e)$ is a result of the following key determinants such as sales process, product reliability, warranty policy, customer report behavior, and the method of discounting.

Suppose that a product with life cycle L is introduced to the marketplace at time zero. This means that the company starts selling the product at time zero and ends up sales at time L . The sales is assumed to occur according to a nonhomogeneous Poisson process $\{N(t); t > 0\}$ with intensity function $\lambda(t)$. Besides, let us define $N(a, b) = N(b) - N(a)$ and $\Lambda(a, b) = \int_a^b \lambda(t)dt$ as the number of units sold and the expected number of sales in the interval $(a, b]$ where $a \leq b$, respectively. It is well proved that the purchase time T of a unit sold during $(a, b]$ is a random variable with density $p(t) = \lambda(t)/\Lambda(a, b)$. Let $f(x)$ and $F(x)$ denote the probability density function and cumulative distribution function for the product's lifetime X , respectively. Suppose a unit failing at age x will be reported to the manufacturer as a warranty claim with probability $1 - q(x)$. This means that the FBNR probability $q(x)$ depends on the failure time of the product. Patankar and Mitra (1995) developed two classes of execution functions, which are $1 - q(x)$, to represent customer report behavior within the warranty period. These two classes of execution functions remain constant one from age zero to a uncertain age and then decreased linearly or exponentially until the warranty expires. These functions capture the authors' notion that customers are more likely to make a warranty claim for an early failure than a late one. Such decreasing report willingness can be attributed to a diminishing utility that the product brings over time. However, the FBNR phenomenon is likely to occur at the very beginning due to unreported defect units, i.e., $q(0) > 0$, which is not captured in their functions. Moreover, their piecewise functions with a random changing point are not smooth enough and will result in serious problems in statistical inference. After carefully examining customer's report behavior, we propose a flexible form for the FBNR function $q(x)$ as

$$q(x) = 1 - r_0 \exp \left[-(ax)^b \right], \quad (1)$$

where $a \geq 0, b > 0$ are scale and shape parameters of $q(x)$, and $r_0 \in [0, 1]$ is the report proportion at time zero. The function (1) is able to describe various report behaviors. For instance, the function may increase slowly at

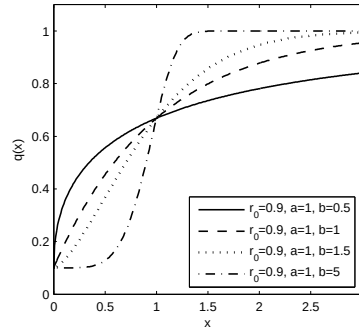


Figure1. FBNR function $q(x)$.

the beginning and then increase fast when the short life cycle is approaching. When $b > 1$ in (1), the function would show the desired trend. For expensive products, the FBNR problem may be rare at the beginning. In this case, we will have $r_0 = 1$. On the other hand, the FBNR function may be constant for a product that are with a long life cycle and are not expensive, i.e., $a = 0$. As can be seen, the FBNR function is flexible enough to capture essential characteristics of a variety of FBNR problems. See Figure 1 for some examples.

We consider a pro-rata rebate warranty policy under which a one-time expense that is not increasing with item lifetime X will be incurred for the item failing within the warranty period. Such policy is often used for non-repairable products that wear out and are not feasibly repaired (Bai and Pham 2004). Examples are automobile tires, car batteries, and television picture tube, etc. Let τ be the warranty length of a product. We define $\alpha(x)$ as the rebate function with a commonly used form given below

$$\alpha(x) = \begin{cases} s(1 - \frac{x}{\tau}), & 0 \leq x \leq \tau \\ 0, & x > \tau \end{cases} \quad (2)$$

where s is the per unit price. Note that the rebate function is not restricted to the above form (2). Generally, it can be a constant, linear, or non-linear function, depending on product warranty terms. For a reported unit sold at t and failing at age x , the discounted warranty cost evaluated at the beginning of study t_s could be obtained as

$$c(x, t) = \begin{cases} \alpha(x)e^{-\delta(x+t-t_s)}, & \text{if } 0 \leq x \leq \tau, t_s \leq x+t \leq t_e \text{ and the failure is reported} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where δ is the continuous discount rate.

3. Discounted Cost Modeling

In this section, we will derive expressions for the mean and variance of the total discounted warranty cost $TDC(t_s, t_e)$ over any period $(t_s, t_e]$ of interest. Essentially, we are interested in all units that are still under warranty as they are likely to fail and lead to warranty expenses within the study period $(t_s, t_e]$. For ease of formulation, let $N(a, b) = N(b) - N(a)$ denote the number of units sold during $(a, b]$ and $DC(a, b)$ denote $N(a, b)$ units' total cost contributions to $TDC(t_s, t_e)$, then we have $DC(a, b) = \sum_{i=1}^{N(a, b)} C_i(a, b)$, where $C_i(a, b)$ is i^{th} unit's cost contribution. As proved in Bai and Pham (2004), if $C_i(a, b), i = 1, 2, \dots, N(a, b)$, are independent variables and have the same first order moment $E[C(a, b)]$ and the second order moment $E[C^2(a, b)]$, then the expected value and variance of $DC(a, b)$ are given by

$$E[DC(a, b)] = \Lambda(a, b)E[C(a, b)], \quad (4)$$

$$Var[DC(a, b)] = \Lambda(a, b)E[C^2(a, b)]. \quad (5)$$

To quantify warranty cost within different periods along the whole time horizon, we show that there are in total nine distinct cases, which are classified based on both starting and ending time of the study. Mean and variance of $TDC(t_s, t_e)$ for each of the nine cases are given below.

Case 1: $0 \leq t_s \leq \tau$ and $t_e \leq \tau$. In this setting, two groups of units are of interest. The first group includes

$N(0, t_s)$ units that are sold during $(0, t_s]$. The warranty expense will be incurred when the units fail and are reported during $(t_s, t_e]$. For a unit with purchase time $t \in (0, t_s]$ and lifetime x , it will lead to a warranty expense only when it fails late, i.e., $x \in (t_s - t, t_e - t]$. Conditional upon the unit sold at time t , the first and second moment of $C(0, t_s)$ are given by $E[C(0, t_s)|T = t] = \int_{t_s-t}^{t_e-t} c(x, t)f(x)[1 - q(x)]dx$ and $E[C^2(0, t_s)|T = t] = \int_{t_s-t}^{t_e-t} c^2(x, t)f(x)[1 - q(x)]dx$. Unconditioning with respect to T , the first and second moment of the cost contribution for a unit sold during $(0, t_s]$ are $E[C(0, t_s)] = \{\int_0^{t_s} \int_{t_s-t}^{t_e-t} c(x, t)f(x)[1 - q(x)]\lambda(t)dxdt\}/\Lambda(0, t_s)$ and $E[C^2(0, t_s)] = \{\int_0^{t_s} \int_{t_s-t}^{t_e-t} c^2(x, t)f(x)[1 - q(x)]\lambda(t)dxdt\}/\Lambda(0, t_s)$, respectively. Thus, $C_i(0, t_s), i = 1, 2, \dots, N(0, t_s)$ have the same first and second order moment. As product lifetime X is independent of its purchase time T and user's report behavior, $C_i(0, t_s)$ are also independent random variables. According to (4) and (5), the mean and variance of $N(0, t_s)$ units' total cost contributions $DC(0, t_s)$ are given by $E[DC(0, t_s)] = \Lambda(0, t_s)E[C(0, t_s)] = \int_0^{t_s} \int_{t_s-t}^{t_e-t} m_1(x, t)dxdt$ and $Var[DC(0, t_s)] = \int_0^{t_s} \int_{t_s-t}^{t_e-t} m_2(x, t)dxdt$, where $m_1(x, t) = c(x, t)f(x)[1 - q(x)]\lambda(t)$ and $m_2(x, t) = c^2(x, t)f(x)[1 - q(x)]\lambda(t)$. The second group consists of the units sold during $(t_s, t_e]$. It is easy to see that $C_i(t_s, t_e), i = 1, 2, \dots, N(t_s, t_e)$ has the same first and second order moment. The expected value and variance of the discounted warranty cost $DC(t_s, t_e)$ incurred are given by $E[DC(t_s, t_e)] = \int_{t_s}^{t_e} \int_0^{t_e-t} m_1(x, t)dxdt$ and $Var[DC(t_s, t_e)] = \int_{t_s}^{t_e} \int_0^{t_e-t} m_2(x, t)dxdt$. The mean and variance of the total discounted cost $TDC(t_s, t_e)$ are given by $E[TDC(t_s, t_e)] = E[DC(0, t_s)] + E[DC(t_s, t_e)] = \int_0^{t_s} \int_{t_s-t}^{t_e-t} m_1(x, t)dxdt + \int_{t_s}^{t_e} \int_0^{t_e-t} m_1(x, t)dxdt$ and $Var[TDC(t_s, t_e)] = Var[DC(0, t_s)] + Var[DC(t_s, t_e)] = \int_0^{t_s} \int_{t_s-t}^{t_e-t} m_2(x, t)dxdt + \int_{t_s}^{t_e} \int_0^{t_e-t} m_2(x, t)dxdt$. After changing the integration order, we will obtain $E[TDC(t_s, t_e)] = \int_0^{t_s} \int_{t_s-x}^{t_e-x} m_1(x, t)dtdx + \int_{t_s}^{t_e} \int_0^{t_e-x} m_1(x, t)dtdx$ and $Var[TDC(t_s, t_e)] = \int_0^{t_s} \int_{t_s-x}^{t_e-x} m_2(x, t)dtdx + \int_{t_s}^{t_e} \int_0^{t_e-x} m_2(x, t)dtdx$.

We show that there are nine cases in total. The remaining eight cases can also be analyzed in a similar fashion. For each of the eight cases, the expected value of the total discounted cost during (t_s, t_e) is given in the following. Note that the variances of $TDC(t_s, t_e)$ for each case can be obtained by replacing $m_1(x, t)$ with $m_2(x, t)$ in the corresponding mean function.

Case 2: $0 \leq t_s \leq \tau$ and $\tau < t_e \leq L$. $E[TDC(t_s, t_e)] = \int_0^{t_s} \int_{t_s-x}^{t_e-x} m_1(x, t)dtdx + \int_{t_s}^{\tau} \int_0^{t_e-x} m_1(x, t)dtdx$.

Case 3: $0 \leq t_s \leq \tau$ and $L < t_e \leq L + \tau$. $E[TDC(t_s, t_e)] = \int_0^{t_s} \int_{t_s-x}^{t_e-x} m_1(x, t)dtdx + \int_{t_s}^{\tau} \int_0^{t_e-x} m_1(x, t)dtdx - \int_0^{t_e-L} \int_L^{t_e-x} m_1(x, t)dtdx$.

Case 4: $0 \leq t_s \leq \tau$ and $t_e > L + \tau$. $E[TDC(t_s, t_e)] = \int_0^{t_s} \int_{t_s-x}^L m_1(x, t)dtdx + \int_{t_s}^{\tau} \int_0^L m_1(x, t)dtdx$.

Case 5: $\tau < t_s \leq L$ and $t_e \leq L$. $E[TDC(t_s, t_e)] = \int_0^{\tau} \int_{t_s-x}^{t_e-x} m_1(x, t)dtdx$.

Case 6: $\tau < t_s \leq L$ and $L < t_e \leq L + \tau$. $E[TDC(t_s, t_e)] = \int_0^{\tau} \int_{t_s-x}^{t_e-x} m_1(x, t)dtdx - \int_0^{t_e-L} \int_L^{t_e-x} m_1(x, t)dtdx$.

Case 7: $\tau < t_s \leq L$ and $t_e > L + \tau$. $E[TDC(t_s, t_e)] = \int_0^{\tau} \int_{t_s-x}^L m_1(x, t)dtdx$.

Case 8: $L < t_s \leq L + \tau$ and $t_e \leq L + \tau$. $E[TDC(t_s, t_e)] = \int_{t_s-L}^{\tau} \int_{t_s-x}^L m_1(x, t)dtdx - \int_{t_e-L}^{\tau} \int_{t_e-x}^L m_1(x, t)dtdx$.

Case 9: $L < t_s \leq L + \tau$ and $t_e > L + \tau$. $E[TDC(t_s, t_e)] = \int_{t_s-L}^{\tau} \int_{t_s-x}^L m_1(x, t)dtdx$.

When $t_s = 0$ and $t_e = L + \tau$, the discounted warranty cost obtained is indeed the discounted life cycle cost, i.e., the present value of overall warranty expense for a product. It can be used to inform the total cost of a warranty program.

4. Special Case

Under special conditions, we can derive closed forms for the mean and variance of the discounted warranty cost. In the following, we provide a detailed discussion for the case 5 (i.e., $\tau < t_s \leq L$ and $t_e \leq L$), meaning that the study period begins after τ periods of time since the product is launched and ends before the product is withdrawn from the market. Other cases can be analyzed in a similar way. Suppose that sales rate is constant over time, the lifetime is exponentially distributed, and $b = 1$ in (1), then we have $\lambda(t) = \lambda$, $f(x) = \eta e^{-\eta x}$, and $1 - q(x) = r_0 e^{-ax}$. If the rebate function $\alpha(x)$ takes the form proposed in (2), meaning that a linear pro-rata policy is adopted, the expectation and variance of the cost are simplified to

$$E[TDC(t_s, t_e)] = \frac{\lambda \eta r_0 s [1 - e^{-\delta(t_e - t_s)}] [\tau(\eta + a) + e^{-\tau(\eta + a)} - 1]}{\delta \tau (\eta + a)^2}, \quad (6)$$

$$Var[TDC(t_s, t_e)] = \frac{\lambda \eta r_0 s^2 [1 - e^{-2\delta(t_e - t_s)}] [(\tau \eta + \tau a - 1)^2 + 1 - 2e^{-\tau(\eta + a)}]}{2\delta \tau^2 (\eta + a)^3} \quad (7)$$

If $\alpha(x) = s$, meaning that a lump sum rebate plan where customers receive a fixed amount of money for a product failure is considered, then the expectation and variance are simply given by

$$E[TDC(t_s, t_e)] = \frac{\lambda \eta r_0 s \left[1 - e^{-\delta(t_e - t_s)} \right] \left[1 - e^{-\tau(\eta + a)} \right]}{\delta(\eta + a)}, \quad (8)$$

$$Var[TDC(t_s, t_e)] = \frac{\lambda \eta r_0 s^2 \left[1 - e^{-2\delta(t_e - t_s)} \right] \left[1 - e^{-\tau(\eta + a)} \right]}{2\delta(\eta + a)}. \quad (9)$$

When $q(x) = 1 - e^{-ax}$, the mean and variance of the discounted cost with the decreasing and constant rebate function could be obtained by replacing r_0 in the above equations with one. Assume that the FBNR function is constant, i.e., $q(x) = 1 - r_0$, the mean and variance of the discounted cost for the time dependent and independent $\alpha(x)$ are given by replacing a in the above equations with zero.

5. Numerical Application

In this section, we will examine impacts of various input variables such as FBNR proportion and warranty length on the discounted warranty cost. We will also discuss how to find the optimal amount of warranty reserve. In warranty reserve management, managers need to set aside certain amount of money for warranty obligations over a given period of time (t_s, t_e) . As discussed above, warranty expense is uncertainty over time because of the randomness within purchase time, product failure time, customer report behavior and other factors. When R amount of money reserved at t_s is smaller than the discounted warranty expense $TDC(t_s, t_e)$, the company needs to find some money from other sources at a much higher cost. Without loss of generality, we assume that the interest rate of raising emergent funding is k ($k > \delta$). Let $L(R, TDC)$ denote the loss of insufficient reserve, it is easy to know that $L(R, TDC) = (TDC - R)[e^{(k-\delta)(t_e - t_s)} - 1]$, where $TDC = TDC(t_s, t_e)$. On the other hand, when the amount of money reserved is larger than the actual expense incurred within the planning period, the company suffers loss of potential gain from other more profitable investment. Let γ denote the rate of return. The loss of excess reserve is given by $L(R, TDC) = (R - TDC)[e^{(\gamma-\delta)(t_e - t_s)} - 1]$. The expected loss is then given by $L(R) = E[L(R, TDC)] = \int_0^R (R - c)[e^{(\gamma-\delta)(t_e - t_s)} - 1]f_{TDC}(c)dc + \int_R^\infty (c - R)[e^{(k-\delta)(t_e - t_s)} - 1]f_{TDC}(c)dc$. We argue that the optimal level of warranty reserve is the value that minimizing the expected loss, i.e., $R^* = \arg \max_R L(R)$. To determine the optimal amount of warranty reserve, the probability distribution of TDC should be specified, which is, however, difficult to obtain. Following the common practice in the warranty reserve literature (Ja et al. 2002; Bai and Pham 2004; Xie and Ye 2016), we simplify the problem by assuming that the probability distribution approximates a normal distribution with mean $E[TDC(t_s, t_e)]$ and variance $Var[TDC(t_s, t_e)]$ that are given in the Section 3. Therefore, we have

$$f_{TDC}(c) = \phi \left[\frac{c - E[TDC(t_s, t_e)]}{\sqrt{Var[TDC(t_s, t_e)]}} \right], \quad (10)$$

where $\phi(\cdot)$ is probability density function of the standard normal distribution with mean zero and variance one.

Without loss of generality, we consider case five, i.e., $\tau < t_s \leq L$ and $t_e \leq L$. Specifically, we assume that $t_e - t_s = 3$. For purpose of illustration, we assume that both sales rate and FBNR function are constant with $\lambda(t) = \lambda = 1000$ and $q(x) = 1 - r_0$. Besides, the lifetime distribution is assumed to be exponential with rate $\eta = 0.05$. We also assume that the rebate function takes general form (2) with $s = 1$. Under the above setting, mean and variance of $TDC(t_s, t_e)$ could be obtained by replacing a in (6) and (7) with zero. We use $k = 10\%$, $\gamma = 6\%$, and $\delta = 3\%$ in the simulation. To examine the effect of customer report behavior and warranty length, the simulation is repeated under various combinations of $q(x)$ and τ .

Table 1 shows mean and variance of the discounted warranty cost together with the optimal level of warranty reserve and the expected loss. As can be seen, the mean and variance of the cost increase as we decrease the FBNR proportion or increase the length of warranty period. Similarly, the table shows more money should be reserved when the expected warranty cost is high and vice versa. These results are consistent with the fact that warranty expense would be reduced when more customers report product failures or when manufacturers shorten the warranty coverage period. Another important observation is that when the anticipated FBNR pro-

Table 1: Mean and variance of the discounted warranty cost together with the optimal level of warranty reserve and the expected loss.

$q(x)$	$\tau = 2$				$\tau = 2.5$			
	$E[TDC]$	$Var[TDC]$	R	$L(R)$	$E[TDC]$	$Var[TDC]$	R	$L(R)$
0.10	124.91	80.35	129.94	1.00	154.86	99.82	160.47	1.12
0.15	117.97	75.88	122.86	0.97	146.26	94.27	151.71	1.08
0.20	111.03	71.42	115.77	0.94	137.65	88.73	142.94	1.05
0.25	104.09	66.96	108.68	0.91	129.05	83.18	134.17	1.02
0.30	97.15	62.49	101.59	0.88	120.45	77.64	125.39	0.98

portion is high, manufacturers could slightly extend the warranty period without causing additional cost. Such longer warranty period is a strong indicator of product's good quality, which may help improve company's reputation as well as boost sales.

6. More Consideration

The knowledge of mean and variance of discounted warranty cost during any period of time helps manufacturers to determine an appropriate level of warranty reserve. In this section, we will briefly discuss the statistical method for the estimation of product lifetime distribution and customer FBNR function that are needed to calculate the mean and variance of the warranty cost. A more comprehensive study on statistical analysis of field return data with FBNR events will be discussed thoroughly in future research.

Essentially, two sets of data are needed in the analysis. The primary data set is the warranty claim data from sold units in which FBNR phenomenon may exist. The data observed from an untracked unit with lifetime X is denoted as (Z, δ) , where δ is the failure-report indicator ($\delta = 1$ if the unit fails before the censoring time C and the failure is reported and $\delta = 0$ otherwise), and Z is the observed service time for the unit, i.e., $Z = \delta X + (1 - \delta)C$. Suppose there are n untracked units and n realizations of (Z, δ) . The observed warranty claim data are then denoted as $\mathbf{D}_{wc} = \{(z_i, \delta_i); i = 1, 2, \dots, n\}$.

The second data set is named as the tracking data from a subset of sold units that are placed under a tracking study and are followed up closely. The FBNR problem in the tracking data is minimal and thus the data can be used to verify the assumption on the lifetime distribution and to decouple the FBNR information from the warranty claim data. The data observed from a tracking unit is $(Z, \tilde{\delta})$, where $\tilde{\delta}$ be the failure-report indicator. Since there is no FBNR problem in the tracking data, $\tilde{\delta}$ is equivalent to the censoring indicator, i.e., $\tilde{\delta} = I_{\{X \leq C\}}$, where I is the indicator function. Suppose there are m tracked units. The observed tracking data are $\mathbf{D}_{tr} = \{(z_j, \tilde{\delta}_j); j = 1, 2, \dots, m\}$.

Let θ_X and θ_q be the parameters in the lifetime distribution and in the FBNR function, respectively. Given $\mathbf{D}_{wc}$ and $\mathbf{D}_{tr}$, the log-likelihood function for $\theta = (\theta_X, \theta_q)$ is

$$\ell(\theta; \mathbf{D}_{wc}, \mathbf{D}_{tr}) = \ell_{wc}(\theta; \mathbf{D}_{wc}) + \ell_{tr}(\theta_X; \mathbf{D}_{tr}), \quad (11)$$

where

$$\ell_{wc}(\theta; \mathbf{D}_{wc}) = \sum_{i=1}^n \left\{ \delta_i \log\{[1 - q(z_i)]f(z_i)\} + (1 - \delta_i) \log \left\{ 1 - \int_0^{z_i} [1 - q(x)]f(x)dx \right\} \right\}, \quad (12)$$

and

$$\ell_{tr}(\theta_X; \mathbf{D}_{tr}) = \sum_{j=1}^m \left\{ \tilde{\delta}_j \log[f(z_j)] + (1 - \tilde{\delta}_j) \log[1 - F(z_j)] \right\}. \quad (13)$$

The ML estimates $\hat{\theta} = (\hat{\theta}_X, \hat{\theta}_q)$ can be obtained by numerically maximizing the joint log-likelihood function

(11). These estimates could then be used in the discounted warranty cost model and used to obtain estimates of warranty reserve.

7. Conclusion

The knowledge of the discounted warranty cost over a specific period of time is essential for manufacturers in warranty reserve management. Customers report behavior significantly affects warranty expense incurred. In this paper, we concentrated on evaluating the discounted warranty cost during any given period of time when the FBNR phenomenon exists. Both mean and variance of the discounted warranty cost for pro-rata policy were derived. We proposed an optimal strategy for warranty reservation and investigated the effect of warranty length and FBNR proportion through a numerical application. The application results show that if ignoring the FBNR behavior, companies will over estimate warranty cost and over reserve money for warranty claims. Our findings can help manufacturers determine the optimal level of warranty reserve for any period of time, which ensures quality of the warranty service as well as the capital efficiency of the company.

References

- Amato, H. N. and Anderson, E. (1976). Note-determination of warranty reserves: An extension. *Management Science*, 22(12):1391–1394.
- Bai, J. and Pham, H. (2004). Discounted warranty cost of minimally repaired series systems. *IEEE Transactions on Reliability*, 53(1):37–42.
- Blischke, W. R., Karim, M. R., and Murthy, D. N. P. (2011). *Warranty Data Collection and Analysis*. Springer.
- Gurgur, C. Z. (2011). Dynamic cash management of warranty reserves. *The Engineering Economist*, 56(1):1–27.
- Ja, S.-S., Kulkarni, V. G., Mitra, A., and Patankar, J. G. (2002). Warranty reserves for nonstationary sales processes. *Naval Research Logistics*, 49(5):499–513.
- Menke, W. W. (1969). Determination of warranty reserves. *Management Science*, 15(10):B–542.
- Patankar, J. G. and Mitra, A. (1995). Effects of warranty execution on warranty reserve costs. *Management Science*, 41(3):395–400.
- Polatoglu, H. and Sahin, I. (1998). Probability distributions of cost, revenue and profit over a warranty cycle. *European Journal of Operational Research*, 108(1):170–183.
- Ulrich, K. T., Eppinger, S. D., et al. (1995). *Product Design and Development*. McGraw-Hill New York.
- Wu, S. (2011). Warranty claim analysis considering human factors. *Reliability Engineering & System Safety*, 96(1):131–138.
- Xie, W. and Ye, Z.-S. (2016). Aggregate discounted warranty cost forecast for a new product considering stochastic sales. *IEEE Transactions on Reliability*, 65(1):486–497.

On assumptions in optimisation of warranty policies

S. Wu*, R. Gitzel** and S. Turrin**

*Kent Business School, University of Kent, Canterbury CT2 7PE, UK. email address: s.m.wu@kent.ac.uk

** ABB AG, Forschungszentrum, Wallstadter Straße 59, 68526 Ladenburg, Germany

Abstract. Optimisation of warranty policy has been a hot research topic in both operations research and statistics communities since warranty providers hope to balance cost-benefit analysis in the nowadays competitive market. Some assumptions are inevitably needed for such research. Most of the existing publications, however, make assumptions that may not be true in practice, based on which biased decision may be made. This paper discusses pitfalls in the assumptions, which include causes of warranty claims, pattern of warranty claims, warranty claim models, field reliability vs product reliability, the relationship between usage and age in 2-dimensional warranty. A real-world example is used to elaborate the arguments.

1. Introduction

Warranty is a duty attached to a product and requires manufacturers to offer pre-specified compensation to buyers when a product fails to perform its designed functions under normal use within the warranty period. Nowadays, product warranty becomes increasingly important in consumer and commercial transactions, and is widely used and serves many purposes (Wu 2013). For example, Congress in the USA passed the Magnusson Moss Act, the European Union (EU) passed legislation requiring all a two-year warranty for all products, and in the UK the annual sales volume of extended warranties is more than £1 billion.

Warranty expense is one of the operating expenses for manufacturers. A product might be sold with a warranty agreement and the manufacturer needs to cover labour and parts needed for repairs or replacement within the warranty period. As a consequence, warranty incurs tremendous cost in the manufacturing industries. For example, the automotive industry spends roughly \$10–\$13 billion per year in the U.S. on warranty claims and up to \$40 billion globally (MSX International Inc 2010).

Warranty policy, which concerns with warranty period and optimisation of maintenance policy of a product during its warranty period, is made based on the estimated reliability of the product and its associated costs. The success of a warranty policy hinges heavily on the well-orchestrated interplay of the good reliability models, cost information, and the effective use of these models.

In the literature, there is considerable research on warranty optimisation (see (Su, Wang 2016, Shafiee, Chukova 2013), for example), which forms a hot research topic in both operations research and statistics communities since warranty providers hope to balance cost-benefit analysis in the nowadays competitive market. Some assumptions are inevitably needed for such research. Most of the existing publications, however, make assumptions that may not be true in practice, based on which biased decision may be made.

This paper discusses the validity of those assumptions, which include causes of warranty claims, pattern of warranty claims, warranty claim models, field reliability vs product reliability, the relationship between usage and age in 2-dimensional warranty.

2. Assumptions in optimisation of warranty policies

Optimisation of warranty policy relies on efficient and unbiased reliability models and accurate information on relevant costs. Unfortunately, in reality, to obtain such models and cost information is very challenging (see (Ye, Ng 2014), for example), due to various constraints, as discussed in the following.

2.1 Information of cost

Many authors optimise warranty policies based on minimising the expected cost relating to a specific topic such as lifecycle cost or maintenance cost, which is vitally important due to the severe market competitiveness. In doing those, various cost information is required. Cost information used to optimise expected long-run cost per unit time is commonly uncertain in practice. For example, repair cost or replacement cost used to optimise preventive maintenance policy can be hard to estimate. Similar optimisation problems can be burn-in policy optimisation, warranty policy optimisation, and maintenance policy optimisation. Hence, precise cost information and reliability models with good quality are usually

hard to obtain, which makes it hard to use existing approaches. As a result, optimisation can bias decision making and hence huge resources (i.e., capital, raw materials, etc) are wasted. A literature review of lifecycle cost optimisation has provided very little information specifically in this area.

The difficulty of acquiring cost information becomes more challenging for products with long-term warranty. During the long warranty coverage, cost of repair and cost of warranty returns become more uncertain than those products with a short-term warranty period. For example, new failure modes not present in the early phases show up in the warranty scope and have to be considered for an accurate estimate of cost.

2.2 Assumption of the reliability model

The concepts of inherent reliability and field reliability should be distinguished, as many authors have already noticed, say (Murthy, Rausand & Østerås 2008) and Figure 1, for example. Inherent reliability is the reliability of a product before the product is shipped out for sale, whereas field reliability is the reliability that is estimated from data collected from the field.

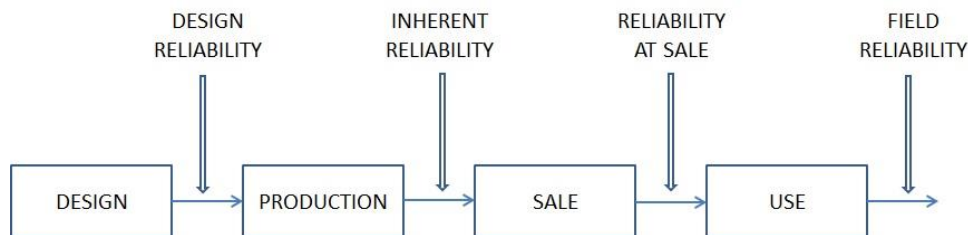


Figure 1. Reliabilities (adapted from {{395 Murthy, DN Prabhakar 2008}})

It should be noted that the field reliability of a product may be significantly different due to its operating environments. For example, a product may have a shorter lifetime in a server weather condition than in a mild weather condition. In particular, situations where the specifications of the product are exceeded (“over-stress”) have a significant impact on the life expectancy of an individual unit. The difference of the inherent reliability of a product may not have such a significant variation. That is, in many cases, to estimate field reliability is not easy.

To optimise the warranty policy for a product, one needs to know the reliability of the product. Normally, warranty policy is optimised before the product is shipped out to its retailers and then used in the field. As a result, optimisation of warranty policies based on inherent reliability may lead to biased results. This is because that many factors influence on warranty claims, as shown in Figure 2 and also more detailed explanation in the following.

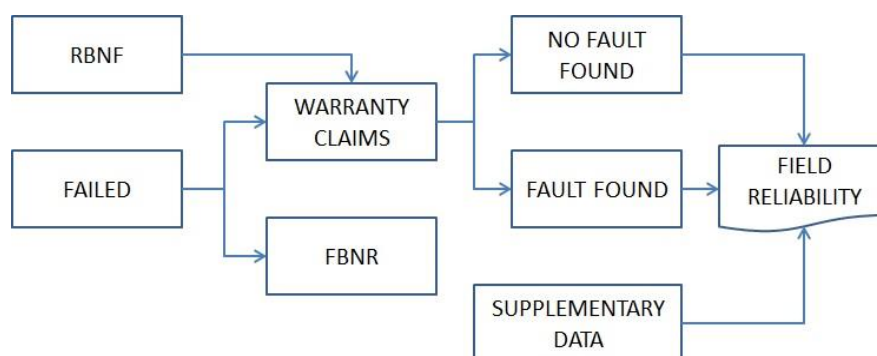


Figure 2. Warranty claim process (RBNF=Reported But Not Failed; FBNR=Failed But Not Reported)

2.2.1 Missing data: Failed But Not Reported (FBNR)

Nowadays, technical systems may also be seen a system composed of users, software systems and hardware systems. It should be noted that product users (e.g., mobile phone users) may not be bothered to claim warranty even if the items are still under warranty coverage due to various reasons. For example, the users might have purchased new and more advanced products (e.g., from Apple iPhone 4 to Apple iPhone 6) or

they might find the effort of claiming warranty too high when compared to the price of the product. This causes a problem of FBNR (failed but not reported). Although there are some research papers attempting to optimise warranty policies with a consideration of FBNR, research on how the field reliability of a product with the existence of FBNR can be estimated is needed. One may of course assume some forms of FBNR functions of time, for example, (Patankar 1996) proposes 4 warranty execution functions. Unfortunately, there are not many research publications which show that those forms of warranty execution functions have been validated with real data, even with data collected from a questionnaire survey. This implies that an extra caution is needed in use of those warranty execution functions.

2.2.2 Unknown failures: no-fault found and Report But Not Failed

Brombacher showed that the observed categories of reliability problems were distributed as: components 21%; customers 17%, apparatus 24% and no fault found 38% (Brombacher 1999). On these statistics, the author further interpreted that the reliability failures in products were split into problems on a component level, problems on “internal product level” (e.g. interaction problems) and problems on a customer/application level. This analysis showed the largest single group where the cause of the failure remained unknown, or no-fault-found (NFF). The NFF phenomenon is a huge challenge when multipart products are dealt with. For example, the NFF contributes on average to 45% of reported service faults in electronic products (Jones, Hayes 2001). The problem is not new, but many believe it is getting worse, in part because today’s highly complex products are equipped with more and more electronic sensors, computers, control functions and wires (Ramsey 2005). Furthermore, for cheap electronic products, the cost of performing a root cause analysis is significantly higher than the worth of the product and thus not regularly performed. This means that the causes of many warranty claims may not be identified and verified.

On the other hand, it has been reported that many claims are fraudulent. This causes another problem of NFBR (not failed but reported). To identify the NFBR phenomenon involves the ability of distinguishing between NFBR and other phenomenon, which may be due to manufacturing problems.

Due to the existence of the NFF and NFBR and their high frequency of occurrence, accurately estimating field reliability becomes a challenge.

2.2.3 Life cycle of a typical product

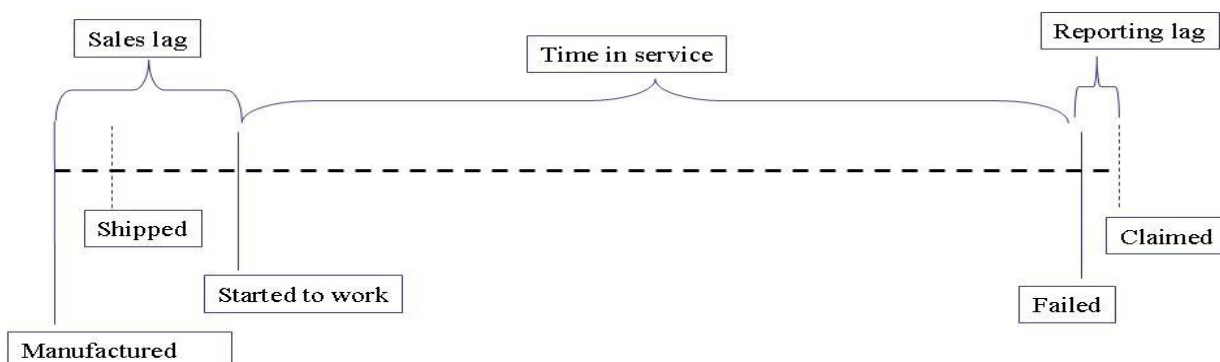


Figure 3. The life cycle of a typical product

The duration of lifetime warranties is defined in the warranty policy with conditions. In general, the term of lifetime is specified by manufacturer as the design life of the product or market life of the product.

Figure 3 shows a typical process from products being manufactured to failed products being received by a manufacturer. The products might be temporarily stored in a warehouse for a time period before they are shipped to distributors or partners or retailers, where the products might spend some time before they are purchased and put in use by users. That is, we can divide the whole period into the following four phases.

- Phase 1 --- This stage starts from the time when the products are manufactured until the time when the products are shipped to warehouses. In this phase, the products are still not shipped to distributors/ partners/retails.

- Phase 2 --- This stage starts from the time when the products are shipped to distributors/partners/retails until the time when they are put in use. There can be several tiers of distributors/partners/retails.
- Phase 3 --- This stage is the period when the products are operated.
- Phase 4 --- Stage 4 starts from the time when products fail until the time the manufacturer has received warranty claims.

Phase 1 and Phase 2 constitute a period of time called *sales lag* and Phase 4 is also called *reporting lag* (see Figure 3). From a reliability analysis perspective, only time in service, or time to failure (TTF), i.e. the time period of Phase 3, is needed but usually it may be difficult to collect the exact time length of Phase 3 because of sales lags and reporting lags may not precisely recorded (See Gitzel et al., 2015, for example).

2.2.4 Design modification and product obsolesce

Due to the competitive nature of the consumer market and customers' demands, manufacturers have started to provide warranties with relatively long period at the beginning of the 21st Century. There are increasing numbers of electrical and electronic product manufacturer who even sell their products with lifetime warranties and some automobile companies have also announced to sell cars with lifetime warranties. Lifetime warranty may persuade customers to buy the products because of its additional assurance which may influence the buyers' intension by convincing them that the products are reliable and can last a long time. Manufacturers who offer lifetime warranty for their products are trying to send the message that they are proud of their products and the customers who bought these products can have a peace of mind for as long as they will keep on using the products. In addition to the difficulty of estimating the cost of lifetime warranty due to the indefinite nature of product lifetime and its service conditions, the manufacturers may also modify product designs and declare products obsolete after they have been in the market for a time period.

Rapid technological advances in many industries, especially in the electronics manufacturing industry, make products obsolete quicker than before, which requires product manufacturers to provide long-term warranties to protect consumers' profits. After products have been declared obsolete, they will no longer be produced. This results in additional complexities as the uncertainties associated with costs incurred by such warranty servicing become more difficult to predict. For example, it is unclear if certain parts or even technologies will be available in the future and if stocked parts will survive storage long enough to be used. Furthermore, the skills required to work with the obsolete product might be lost due to retirements and personnel fluctuation driving up the required person hours and person hour cost. Also, completely new failure modes might occur as materials start to deteriorate in ways not normally considered. All this poses a challenge on optimising warranty servicing policy.

For example, Figure 4 shows the number of warranty claims that had been received. In month 28, a major modification on the product was conducted. As a result, the number of warranty claims became increasing. Of course, a product modification does not necessarily mean that the reliability of the product is also been improved. Such modifications can be from a function perspective. In month 73, the product was announced obsolete. Following that, warranty claims become fewer. For such products, to obtain a precise estimate of the number of warranty claims, or further to estimate the reliability of the products on the basis of warranty claims, is extremely difficult due to the uncertainty presented.

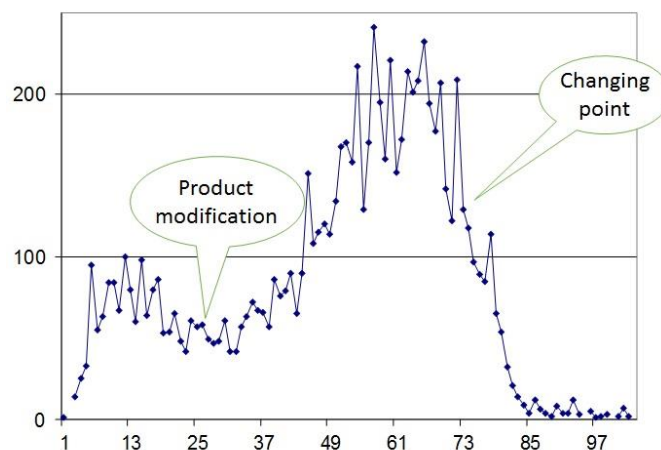


Figure 4. The number of warranty claims of a networking product.

2.3 Assumption on the relationship between age and usage

It can be found that a linear relationship between the age and the accumulated usage of a product (“running hours”) is assumed by many authors. Optimisation of two-dimensional warranty policies needs first to understand the relationship between age and usage (see (Ye et al. 2013)). It should be noted, however, such an assumption may not hold due to the asymmetry existing between the two age and usage. Items with older age may have lower accumulated usage, but items with higher accumulated usage normally have older age (see Wu, 2014). Also, emergency equipment (such as an electrical motor used for emergency shutdown) or spare units might be quite old without having experienced a lot of actual use. As such, the linearity assumption of the relationship between age and usage needs validation before it is made.

2.4 Assumptions on modelling times between claims

In the literature, many reliability datasets have been published. Those datasets normally include data of time-between-failures. However, most of them do not indicate the failure modes, which may cause the problem that survival distributions are estimated based on time-between-failures due to failures of completely different components. Some publications do not check the statistical properties before they estimate models based on different assumptions. For example, a series of time-between-failure data are statistically independent, based on which a non-homogeneous Poisson process is estimated. On the other hand, when a series of time-between-failures are statistically dependent, but based on the data a renewal process or a geometric process is developed. In our experience, this is especially problematic in the case of repairable systems. It is highly problematic to collect the Time Between Failures (TBF) for failures that have nothing to do with each other such as a mechanical and an electronics failure in a robot and consider the robot as good as new afterwards.

2.5 Warranty policies for complex systems

Nowadays, technical systems become increasingly complex. They may be composed of software systems (e.g., embedded software systems for controlling their host hardware systems) and hardware systems. Authors normally suppose that the software and hard systems are structured in series, based on which warranty policies are derived. As a result, they have little difference with those derived on the assumption of single hardware systems. While such an assumption may be correct for some systems, it does not hold in many systems. The failure mechanism of software is different from that of hardware in that software failures are almost exclusively design problems. To make things worse, software tends to evolve over time due to updates, so it is hard to compare units of a system running different versions of the software (and having updated at different points in time). Normally, software systems have different warranty periods with hardware ones and their maintenance mechanisms are also different. As such, there is a need to study warranty management for such complex systems so that optimal warranty policies can be sought.

3. A real-world example

Table 1 shows the time between warranty claims of 22 subsystems of the same type of a product. Some users hope to sign extended warranty contract, which requires the manufacturer to optimise the price and duration of the extended warranty. The failure modes recorded in the warranty claims and maintenance log are on subsystem levels. The cost of repair on the claimed items is not free of noise as the costs on the same failure mode in some cases differ significantly. This suggests that it is unrealistic to derive a precise solution through optimising the expected cost, as normally did in many publications. Apparently, this is a typical example of optimisation of warranty policy for complex systems,

We have tried to fit the data with the renewal process, the geometric process, the non-homogeneous process with the power law (NHPP), and the virtual age process. Based on the AIC values of the processes, the

NHPP-PL is selected. The expected number of failures in the first t hours is $\left(\frac{t}{239.299}\right)^{1.238}$. Given that the product has already survived for u time units, then the expected number of failures within next v time units is $\left(\frac{u+v}{239.299}\right)^{1.238} - \left(\frac{u}{239.299}\right)^{1.238}$. More importantly, a confidence interval of this estimated number of

failures should be given in order to allow for the decision makers to understand the uncertainty associated with this expected number when they determine the duration of extended warranty and further price it. Of course, one may also use the compound Poisson process to model the claim process, with which the uncertainties in both models of time-between-claims and cost on the claims are considered.

Table 1. Times between warranty claims (in day) (where x_1 is the time to first claim, x_k are times between the $(k - 1)$ th and the k th claims, where $k = 2, 3$.)

No.	x_1	x_2	x_3	No.	x_1	x_2	x_3	No.	x_1	x_2	x_3	No.	x_1	x_2	x_3
1	142	56	35	7	530	59	266	13	46	93	41	19	97	41	413
2	201	259	21	8	14	403	223	14	448	45	9	20	436	795	207
3	206	281	29	9	314	15	176	15	378	169	261	21	648	22	175
4	26	195	423	10	60	9	28	16	270	16	103	22	315	231	54
5	251	461	117	11	117	26	156	17	301	38	45				
6	74	269	552	12	146	127	62	18	304	32	42				

4. Conclusion

This paper discussed main challenges in optimisation of warranty policies. Those challenges are mainly due to the reliability and validity of data collected from warranty claims and assumptions needed in warranty policy optimisation.

From the above discussion, it can be seen that there are many sources of uncertainty existing in the process of optimisation of warranty policy. To reduce the uncertainty and mitigate the risk of unbiased decision making, we suggest that more research attention should be paid on novel methods of dealing with risk and uncertainties in warranty policy optimisation. For example, more studies on risk based decision making methods in warranty optimisation are therefore needed.

Acknowledgement

This research was partly supported by the Economic and Social Research Council of the United Kingdom (Project Ref: ES/L011859/1).

References

- Brombacher, A.C. 1999, "Maturity index on reliability: covering non-technical aspects of IEC61508 reliability certification", *Reliability Engineering & System Safety*, **66**(2), 109-120.
- Gitzel, R., Turring, S. and Maczey, S., 2015, July. A Data Quality Dashboard for Reliability Data. In *2015 IEEE 17th Conference on Business Informatics (CBI)*, Vol. 1, pp. 90-97.
- Jones, J. & Hayes, J. 2001, "Investigation of the occurrence of: no-faults-found in electronic equipment", *IEEE Transactions on Reliability*, **50**(3), 289-292.
- MSX International Inc 2010, *Integrated Warranty Solutions—White Paper*.
- Murthy, D.P., Rausand, M. & Østerås, T. 2008, *Product reliability: specification and performance*, Springer Science & Business Media.
- Patankar, J.G. 1996, "Warranty and consumer behavior: Warranty execution" in *Product Warranty Handbook*, eds. W.R. Blischke & D.N.P. Murthy, Marcel Dekker Inc, New York, pp. 421-438.
- Ramsey, J. 2005, "Special report: avoiding NFF", *Avionics today*.
- Shafiee, M. & Chukova, S. 2013, "Maintenance models in warranty: A literature review", *European Journal of Operational Research*, **229**(3), 561-572.
- Su, C. & Wang, X. 2016, "Optimal Upgrade Policy for Used Products Sold with Two-dimensional Warranty", *Quality and Reliability Engineering International*, DOI: 10.1002/qre.1973
- Wu, S. 2013, "A review on coarse warranty data and analysis", *Reliability Engineering and System Safety*, **114**(1), 1-11.
- Wu, S., 2014. Construction of asymmetric copulas and its application in two-dimensional reliability modelling. *European Journal of Operational Research*, **238**(2), pp.476-485.
- Ye, Z., Murthy, D.P., Xie, M. & Tang, L. 2013, "Optimal burn-in for repairable products sold with a two-dimensional warranty", *IIE Transactions* **45**(2), 164-176.
- Ye, Z. & Ng, H.K.T. 2014, "On analysis of incomplete field failure data", *The Annals of Applied Statistics* **8**(3), 1713-1727.

Mission Reliability Evaluation of Repairable PMS with Bayesian Networks

Xiaoyue Wu*

**College of Information Systems and Management, National University of Defense Technology, Changsha, Hunan, 410073, China. Email address: xiaoyuewu@nudt.edu.cn*

Abstract. Because of its importance both in theory and application, reliability evaluation of phased mission systems (PMS) has been investigated widely in literature. Although Bayesian networks (BN) have been used for mission reliability evaluation of PMS, repair activity of system components was not fully considered. In this paper, a Bayesian network based modeling approach is presented for PMS with repairable components. The behavior of component is described by component nodes of different times with conditional probabilities given by continuous time Markov chains (CTMC). The mission reliability is approximated by evaluating the system states at discrete time points based on the success-state sampling approach for system mission reliability. For each discrete time point, a system state node is defined as the parent node of all its component state nodes in accordance with phase success criterion. The system mission reliability is evaluated by using probabilistic reasoning methods. When the number of the discrete time points becomes large enough, the estimated value of mission reliability will converge to the precise one. An example is used for illustration purpose, and results are compared with that of analytical solutions.

1. Introduction

Phased-mission system (PMS) refers to a dynamic system that executes its mission in consecutive phases, and its configuration, operational requirements and the duration of its phase may change from phase to phase (Xing and Amari, 2008). In engineering application, many systems can be modeled as phased-mission systems (PMS), which usually undertake critical missions. For example, the Spaceflight Telemetry, Tracking and Control System (TTC) is a typical PMS, which must provide highly reliable service for spacecrafts during mission time. Since a spacecraft orbits the earth, it can pass over a ground station for a limited duration, which can be seen as a phase for TTC system (Guest, 2013).

During the last three decades, lots of research works have been done on reliability modeling and evaluation of PMS because of its practical importance and theoretical challenge (Xing and Amari, 2008). The existing methods for reliability modeling and evaluation of PMS can be classified into 4 main categories: combinatorial methods, state-space methods, hybrid methods and simulation methods. Combinatorial methods include Boolean algebra, fault trees(FT), binary decision diagrams(BDD) (Xing and Amari, 2008), which are computationally efficient, but are difficult to consider reparability of system components. State-space methods mainly based on continuous time Markov chains (Alam and Al-Saggaf, 1986), which can describe the dynamic behaviors of system components. However, they have state explosion problem when the number of components becomes larger. Simulation methods are flexible in modeling complex systems, but they usually require much more time to get the reliability estimation results with satisfied precision (Zhou et al., 2011). Hybrid methods (also called modular methods) intend to get a balance between computational efficiency and modeling ability (Ou et al. 2002; Wang and Trivedi, 2007). Typically, the dynamic modular (commonly at component level) is described by CTMC, and the static modular (usually at system level or phase level) is modeled by using combinatorial methods such as FT or BDD methods. However, the FT approach has relatively lower computational efficiency, and the BDD approach has complicated modeling process.

Bayesian network has been used for evaluating the mission reliability of PMS (Liu et al., 2008). However, repair activity of components has not been considered in existing literature. In this paper, we present a method for evaluating the mission reliability of PMS with repairable components based on Bayesian network. This method can take the repair activity of the components into consideration, and the modeling process is straightforward. The remaining parts of this paper are organized as follows. Section 2 presents the main assumptions and the modeling process of the Bayesian network for evaluating the mission reliability of PMS. Section 3 gives a numerical example of PMS, the results by analytical method and the method of this paper are compared for verification purpose. Section 4 concludes the works of this paper.

2. Mission Reliability Model

2.1 Assumptions

This paper makes the following assumptions about the system and its components.

- (a) The system has M phases, $ph_1, ph_2, \dots, ph_M$, and totally N components: $cp_1, cp_2, \dots, cp_N$.
- (b) The state of the system considered and the states of its components are binary-valued. Let x_i denote the state of component cp_i , then $x_i = 1$ means cp_i is up, or operational, $x_i = 0$ means that cp_i is down (or under repair).
- (c) The duration of each phase is deterministic. Let T_i stand for the mission duration of ph_i . The state of the system in ph_i , denoted by s_i , is determined by the states of its components via the structure function ϕ_i as given below.

$$s_i = \phi_i(x_1, x_2, \dots, x_N) \quad (1)$$

- (d) The failure and repair events of all components are statistically independent. The time-to-failure and repair time of system components are all exponentially distributed. For component $cp_i, i = 1, \dots, N$, its failure rate and repair rate are λ_i, μ_i respectively.
- (e) Repair work is perfect, component will behavior as new one after being repaired.
- (f) A phase mission success requires that the system remain in up state throughout the entire phase duration without entering into down state. The mission success of PMS requires that the mission success in all of its phases.

2.2 Approximated Estimation of Mission Reliability

For ph_j , its start time is $tb_j = \sum_{i=0}^{j-1} T_i$, and its end time is $te_j = tb_j + T_j$. Let the duration of ph_j be equally divided into L_j intervals, each of which will have length $\delta_j = T_j / L_j$.

$$tb_j = t_{j,0} < t_{j,1} < t_{j,2} < \dots < t_{j,L_j} = te_j.$$

Suppose that at time $t_{j,k}$, the state probability vector of cp_i is $\pi_{j,k}^i = (p_0, p_1)$, where p_0, p_1 is the probability that cp_i is in down state and up state respectively. Let Q_i be the infinitesimal generator matrix of the CTMC associated with cp_i

$$Q_i = \begin{bmatrix} -\mu_i & \mu_i \\ \lambda_i & -\lambda_i \end{bmatrix}. \quad (2)$$

By CTMC theory (Ross, 2010), at time $t_{j,k+1}$, the state probability vector of cp_i will be

$$\pi_{j,k+1}^i = \pi_{j,k}^i \exp(Q_i \delta_j) = \pi_{j,k}^i E^{i,j}. \quad (3)$$

where for late use, we define $E^{i,j}$ as

$$E^{i,j} = \exp(Q_i \delta_j) = \begin{bmatrix} e_{0,0} & e_{0,1} \\ e_{1,0} & e_{1,1} \end{bmatrix}. \quad (4)$$

Let $U_{j,k}$ be the event that the system is in up state at time $t_{j,k}$, and $O_{j,k}$ represent the event that the system never enters into failure state during $[0, t_{j,k}]$. Based on the ideas of the success-state sampling method by Lu et al. (2015), we can see that the occurrence of $O_{j,k}$ implies that until the time $t_{j,k}$, the system must stay in operational state at every time point no later than $t_{j,k}$. i.e.

$$O_{j,k} = \lim_{\delta \rightarrow 0} \bigcap_{\alpha \leq j} \bigcap_{\substack{t_{\alpha,\beta} \leq t_{j,k} \\ \beta \in \{1, \dots, L_j\}}} U_{\alpha,\beta} \approx \bigcap_{\alpha \leq j} \bigcap_{\substack{t_{\alpha,\beta} \leq t_{j,k} \\ \beta \in \{1, \dots, L_j\}}} U_{\alpha,\beta}. \quad (5)$$

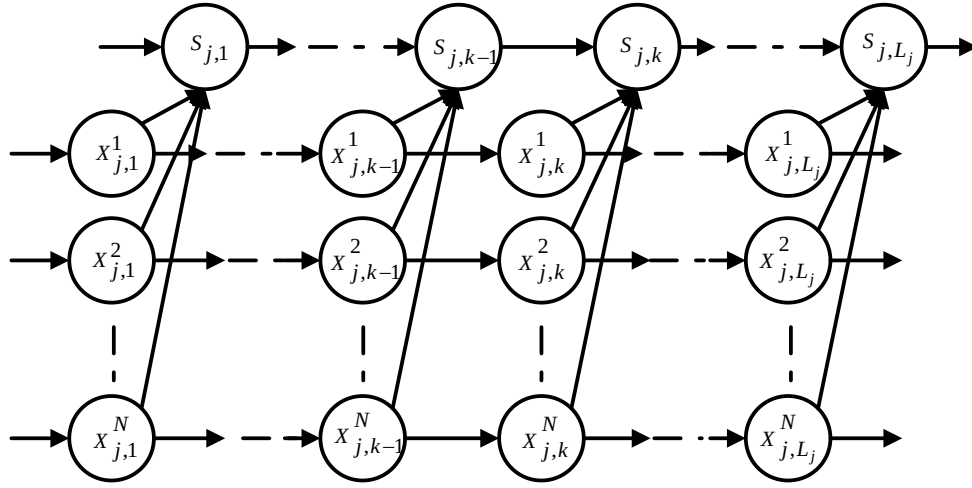


Figure 1. Bayesian network model for a phase.

2.3 System State Model

Let $X_{j,k}^i$ stand for the state of cp_i at the time point $t_{j,k}$, and $S_{j,k}$ denote the state of the system at the time point $t_{j,k}$. Then by (1) we have

$$S_{j,k} = \phi_j(X_{j,k}^1, X_{j,k}^2, \dots, X_{j,k}^N) \quad (6)$$

Fig. 1 is the Bayesian network that describes the behavior of the PMS in ph_j .

For a given phase ph_j , because that the different components are mutually statistically independent, the states of the component at time $t_{j,k}$ is influenced by its state at previous time $t_{j,k-1}$. Therefore, there is a directed edge from $X_{j,k-1}^i$ to $X_{j,k}^i$, namely, the parent node of $X_{j,k}^i$ is $X_{j,k-1}^i$. By (4), the conditional probability table (CPT) for node $X_{j,k}^i$ can be given by

$$\Pr\{X_{j,k}^i = 1 | X_{j,k-1}^i = \gamma\} = e_{\gamma,1}, \quad \gamma = 0, 1, \quad \forall k = 2, 3, \dots, L_j \quad (7)$$

Recall (5) and the requirement that if the mission success in ph_i , the system state at any time point within the phase must be in up state, and the system at $t_{j,k}$ is in up state only if it is in up state at all its previous time points. So, node $S_{j,k-1}$ is introduced as the parent node of node $S_{j,k}$.

Since the state of the system is determined by the states of its components, and node $S_{j,k}$ is used to represent the state of the system at time $t_{j,k}$. The CPT of $S_{j,k}$ can be given by (6) as below

$$\Pr\{S_{j,k} = 1 | S_{j,k-1}, X_{j,k}^1, \dots, X_{j,k}^N\} = \begin{cases} 1 & \text{if } \phi_j(X_{j,k}^1, X_{j,k}^2, \dots, X_{j,k}^N) = 1, S_{j,k-1} = 1 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

2.4 Mission State Model

According to the assumptions of PMS, the state of a component at the beginning time of ph_j is its state at the end time of the previous phase of ph_{j-1} . Thus, as shown in Fig. 2, the Bayesian networks of all phases of the PMS can be connected to form an integrated Bayesian network of the PMS in the following way.

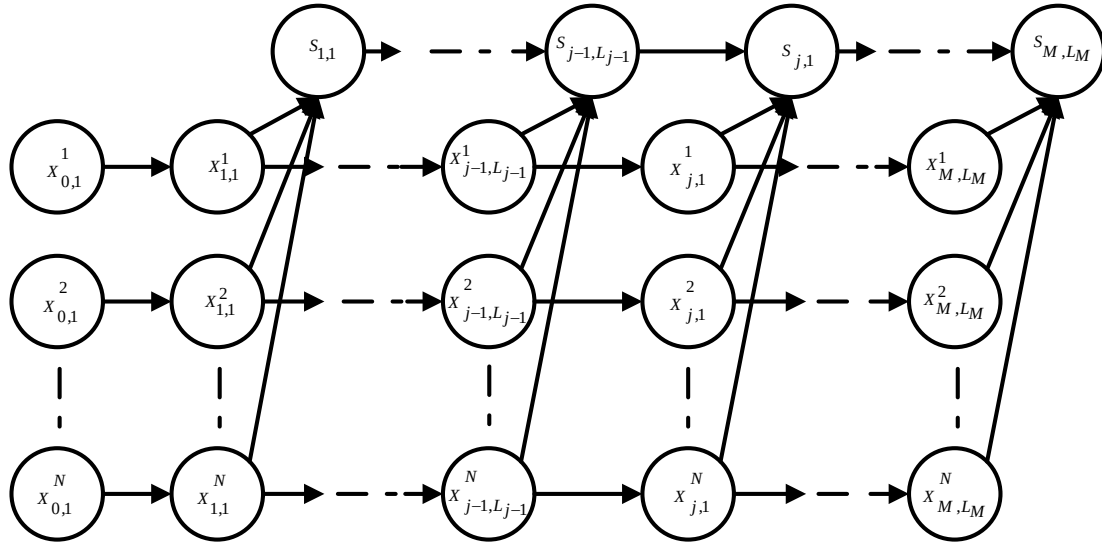


Figure 2. Bayesian network model for PMS

- (a) Set $X_{j-1,L_{j-1}}^i$ as the parent node of $X_{j,1}^i$, $\forall i=1, \dots, N, j=2, \dots, M$.
- (b) Set $S_{j-1,L_{j-1}}$ as the parent node of $S_{j,1}$, $\forall j=2, \dots, M$.
- (c) Introduce binary valued node $X_{0,1}^i$ to represent the initial state of cp_i , and set it as the parent node of $X_{1,1}^i$, $\forall i=2, \dots, N$.

For mission success of the PMS, the state at the end time of the last phase must be in up state. Because we have assumed that at the beginning time all components are in up state, the mission reliability of the PMS can be evaluated by making the probability inference about query variable S_{M,L_M} with evidence $\{X_{0,1}^i = 1, i=1, 2, \dots, N\}$ as below

$$R = \Pr\{S_{M,L_M} = 1 \mid X_{0,1}^1 = 1, X_{0,1}^2 = 1, \dots, X_{0,1}^N = 1\} \quad (9)$$

The above probability inference can be easily solved by using inference algorithms of Bayesian networks, such as variable elimination and likelihood weighted sampling algorithms (Koller, 2009).

3. Numerical Example

Suppose we have a PMS with two components, and two phases. The distribution parameters of the two components are $\lambda_1 = 1/50$, $\lambda_2 = 1/50$, $\mu_1 = 1/2$, $\mu_2 = 1/2$. The structure functions, durations of the two phases are assumed to be $\phi_1(x_1, x_2) = x_1 + x_2$, $\phi_2(x_1, x_2) = x_1 \cdot x_2$ and $T_1 = 3, T_2 = 3$. Thus, the generator matrices of the two phases can be written respectively as

$$Q_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \lambda_2 & -\lambda_2 - \mu_1 & 0 & \mu_1 \\ \lambda_1 & 0 & -\lambda_1 - \mu_2 & \mu_2 \\ 0 & \lambda_1 & \lambda_2 & -\lambda_1 - \lambda_2 \end{bmatrix}$$

$$Q_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & \lambda_2 & -\lambda_1 - \lambda_2 \end{bmatrix}$$

Table 1. The mission reliabilities with different numbers of intervals

L_1	5	10	30	50	100	150	180	300	450
L_2	5	10	30	50	100	150	180	300	450
R	0.8611	0.8480	0.8385	0.8365	0.8350	0.8345	0.8343	0.8340	0.8338

Assume that at the beginning time, the two components are both in up state. By using the traditional CTMC model of PMS (Alam and Al-Saggaf, 1986), we can obtain the mission reliability of this PMS as $R = 0.833476$.

To use the proposed method for this example, a Bayesian network model is built with $N = 2$, $M = 2$. The model is solved by using variable elimination algorithm (Koller and Friedman, 2009). Table 1 presents the estimated mission reliabilities for different numbers of intervals. From Table 1, we can see that the estimated mission reliability of the PMS converges to the reliability value obtained by using CTMC when the number of intervals increases.

4. Conclusion

By using Bayesian networks, this paper present a new hybrid method for modeling and evaluation of mission reliability of repairable PMS. The reparability of the system components can be considered in a natural way. The modeling process is straightforward. Although the evaluation is based on discretization approximation in time horizon, the accuracy of the estimated value can be improved by decreasing the length of intervals. Moreover, the model can be solved by using existing Bayesian network tools. In the future, the proposed modeling method will be extended for evaluating the reliability of more complex PMS.

References

- Alam M, Al-Saggaf UM (1986) Quantitative reliability evaluation of repairable phased-mission systems using Markov approach. *IEEE Transactions on Reliability*, 35(5):498-503.
- Guest AN (2013) Telemetry, Tracking, and Command (TT&C), in *Handbook of Satellite Applications*, volume 1, Springer, pp. 1067-1078
- Koller D, Friedman N (2009) *Probabilistic Graphical Models: Principles and Techniques*. Adaptive Computation and Machine Learning Series. The MIT Press
- Liu D, Zhang CY, Xing WY, Li R (2008) Reliability analysis of phased-mission systems using Bayesian networks In *Proc. 2008 Annual Reliability and Maintainability Symposium*, pp. 21-26.
- Lu JM, Wu XY, Liu YL, Lundteigen MA (2015) Reliability analysis of large phased-mission systems with repairable components based on success-state sampling. *Reliability Engineering & System Safety*, 142:123-133.
- Ou Y, Meshkat L, Dugan BJ (2002) Multi-phase reliability analysis for dynamic and static phases. In *Proc. 2002 Annual Reliability and Maintainability Symposium* pp. 404 -410.
- Ross SM (2010) *Introduction to Probability Models*, 10th Edition. Academic Press.
- Wang DZ, Trivedi KS (2007) Reliability analysis of phased-mission system with independent component repairs. *IEEE Transactions on Reliability*, 56(3): 540-551.
- Xing LD, Amari SV (2008) Reliability of Phased-mission Systems. In *Handbook of Performability Engineering* (Krishna B. Misra Ed.), Springer, pp349-368
- Zhou H, Li F, Wei H (2011) A Simulation Model for Multi-platform Phased Mission System Reliability Analysis. In *Proc. 2011 9th International Conference on Reliability, Maintainability and Safety (ICRMS)*, Academic Press, Boston, pp.151-158.

Inspection and Maintenance Optimization of a Multi-state Series – parallel System

Ming-Ang YIN*, Zhi-Li SUN**, Jian WANG*** and Yu GUO****

* School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110004, China
yinma@me.neu.edu.cn

** School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110004, China
zhlsun@mail.neu.edu.cn

*** School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110004, China
jianwang8805@163.com

**** School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110004, China
guoyu@me.neu.edu.cn

Abstract. This paper presents a structure, inspection and maintenance optimization method for a multi-state series-parallel system considering economic dependence and state-based inspection and maintenance. The objective function in this paper is the sum of the average maintenance cost per unit time and the system installed cost. The constraint function is the system availability. The semi-regeneration theory and the universal generating function (UGF) are applied to evaluate objective function and availability. A hybrid algorithm of particle swarm optimization (PSO) and local search (LS) is proposed to design the system structure and maintenance strategy with the minimum cost to achieve the desired level of availability. Cases show that the state-based maintenance combined with system structure can not only save cost, but also get optimal solution.

1. Introduction

Loss of halting production increases with the improvement of its high level of large-scale, high performance and automation. The quality, output and total economic benefit of product will be more and more dependent on units, which lays stress on their reliability. Effective maintenance strategy is an important method to improve its reliability. Before they solve such a combinational optimization problem of units during producing process, engineers estimate the version and amount of units according to engineering experience on the base of some former and typical design case and then organize production. Constrained by the lack of engineering cases and complex configuration of system, maintenance strategy of units is only to be described qualitatively. Optimizing version, amount of units and relevant maintenance strategy is the problem to be solved under the constraint of both highest productivity and lowest running cost of production line. So, configuration of a production line belongs to a combinational optimization problem that configuration and maintenance strategy of units in each process (Kuo & Wan, 2007; Shafiee & Chukova, 2013; Kuo & Zuo, 2003).

Literature (Abdelbar & Abdelshahid, 2004; Clerc & Kennedy, 2002; Nahas, Khatab, Ait-Kadi, & Nourelfath, 2008; Park, Lee, Shin, & Lee, 2005; Yan, Shih, & Lee, 2011) declared that Meta-heuristic Algorithms such as Genetic Algorithm(GA), Tabu Search(TS) and Ant Colony Optimization(ACO) were used presently to optimize multi-state series-parallel system, literature (Kuo & Wan, 2007) reviewed some former research on algorithms and proposes that there was still a chance to enhance optimization efficiency by the usage of other algorithms, especially hybrid algorithm. Hybrid algorithm of Particle Swarm Optimisation (PSO) and Local Search(LS) was applied in this paper. Literature (Abdelbar & Abdelshahid, 2004; Yan, et al., 2011) performed initial research on such a hybrid algorithm.

This paper mainly focuses on installation and maintenance cost are seen as the objective function, while availability of the system is seen as the constraint function, PSO combined with LS algorithm is applied to optimize such a problem, and a novel local search method is put forward.

2. Description and Hypothesis of Problem

2.1 Description

In a typical multi-state series-parallel system, N subsystems are connected in series, each of them is consisted of $M_i (i \in \{1, 2, \dots, N\})$ same units in parallel. Configuration of one production line is shown in Fig.1. With regard to the i th subsystem, there are V_i different versions available in market. To make things simplified, the i th subsystem in which units with version of j (where $j=1, 2, \dots, V_i$) is applied as an example below. $C_{IN,ij}$ represents installation cost. There are K_{ij} states in total for the unit from brand new to failure. State transition probability matrix of the unit is $\mathbf{P}_{ij} \in R^{K_{ij} \times K_{ij}}$. $X_{ijl}(t)$ represents the state of l th (where $l=1, 2, \dots, M_i$) unit at moment t . $X_{ijl}(t)=1$ represents that the l th unit is brand new at moment t and $X_{ijl}(t)=K_{ij}$ represents that this unit is failure. Productivity of the unit will be $\gamma_{ij}(w)$ if $X_{ijl}(t)=w$, while productivity of invalid units will be zero, that is $\gamma_{ij}(K_{ij})=0$. Productivity of subsystem is the sum of that of all its components, while productivity of the whole system is the minimum of that of all its subsystems.

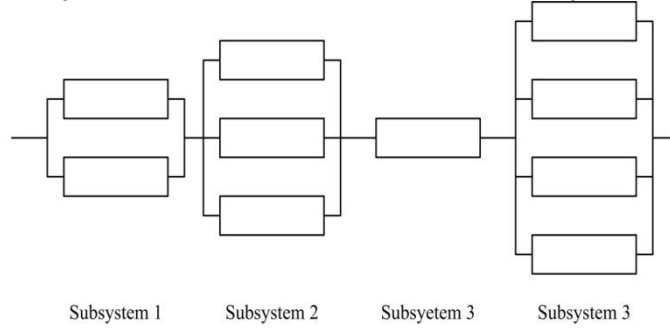


Fig.1. Schematic diagram of a typical series-parallel system

To guarantee that the system runs normally, Maintenance method of subsystem can be judged by the worst state of unit in the subsystem $X_{ij1}^{(w)}$. To avoid unnecessary inspection, multi-inspection threshold $L_{INS,d,ij}$ will be applied, where $d=1, 2, \dots, N_{INS,ij}$. $N_{INS,ij}$ represents the number of inspection threshold of the subsystem, and furthermore, the maximum inspection interval is $N_{INS,ij}+1$ unit time. For example the system will be inspected after d unit time when $L_{INS,d-1,ij} > X_{ij1}^{(w)} \geq L_{INS,d,ij}$. Maintenance method of such a unit will be determined according to preventive maintenance threshold $L_{PM,ij}$ and opportunistic maintenance threshold $L_{OM,ij}$. When $L_{PM,ij} \leq X_{ij1}^{(w)}$, all the failure units will be replaced, while other units of which the states exceed $L_{OM,ij}$ will be maintained preventively. Maintenance strategy of subsystem i can be expressed as

$$\theta_{ij} = [L_{PM,ij}, L_{OM,ij}, L_{INS,1,ij}, L_{INS,2,ij}, \dots, L_{INS,N_{INS,ij},ij}]^T \quad (1)$$

Where

$$K_{ij} \geq L_{PM,ij} \geq L_{OM,ij} \geq L_{INS,1,ij} \geq L_{INS,2,ij} \geq \dots \geq L_{INS,N_{INS,ij},ij} \geq 1$$

2.2 Hypotheses

Some important hypotheses used in this topic are listed below.

(1) Ability of units degrades with time goes, such a degradation process follows discrete time, discrete state Markov process.

(2) Inspection and maintenance time can be ignored compared to working time.

(3) The units recover to be brand new after maintenance that is only perfect maintenance exists.

(4) Failure units can only be inspected when it's being tested, which can be called 'recessive failure'.

3. Calculation model

3.1 Installation cost and maintenance cost of the system

Installation cost and maintenance cost of the system can be divided into two parts. One is the installing cost of the system. The other is maintenance cost of the system where mean maintenance cost during semi-regeneration period is applied.

Installation cost C_{IN} of system can be expressed as

$$C_{IN} = \sum_{i=1}^N M_{ij_i} \times C_{IN,ij_i} \quad (2)$$

Maintenance scheme will be confirmed according to inspection results after inspection, furthermore, according to the hypothesis (2) that maintenance is instantaneous, so maintenance of the system occurs at inspection point, mean inspection and maintenance $\bar{C}_{M,ij}(\theta_{ij})$ of subsystem in unit time can be calculated through renewal theory.

$$\bar{C}_{M,ij} = (C_{INS,ij} + E_{CH,ij}(\theta_{ij}) + E_{CM,ij}(\theta_{ij}) + E_{CPM,ij}(\theta_{ij})) / E_{TSR,ij}(\theta_{ij}) \quad (3)$$

Where

$$E_{TSR,ij}(\theta_{ij}) = \sum_{w=1}^{S_{ij}} \left(\left(\pi_{ij}^{SR}(\theta_{ij}) \right)_w \Delta_{INS,ij}(MA, \theta_{ij}) \right) \quad (4)$$

represents the length of semi-regeneration cycle of the i th subsystem under θ_{ij} . $E_{CH,ij}(\theta_{ij})$, $E_{CCM,ij}(\theta_{ij})$ and $E_{CPM,ij}(\theta_{ij})$ represent the expectation cost of halting, corrective and preventive maintenance of the subsystem during semi-regeneration cycle separately, where preventive maintenance cost also includes opportunistic maintenance cost. All maintenance occurs at semi-regeneration point, so all the above derivation about cost are based on the stable state probability vector $\pi_{ij}^{SR}(\theta_{ij})$ (Zhou YF, & Zhang ZS, 2013). Expectation of halting cost of subsystem during semi-regeneration cycle can be expressed as

$$E_{CH,ij}(\theta_{ij}) = C_{H,ij} \sum_{w=1}^{S_{ij}} \left(\left(\pi_{ij}^{SR}(\theta_{ij}) \right)_w I(X_{ij,1}^{(w)} \geq L_{PM,ij}) \right) \quad (5)$$

All failure units should only be replaced after being tested. So expectation of corrective cost of subsystem can be expressed as

$$E_{CCM,ij}(\theta_{ij}) = C_{CM,ij} \sum_{w=1}^{S_{ij}} \left(\left(\pi_{ij}^{SR}(\theta_{ij}) \right)_w \sum_{l=1}^{M_i} I(X_{ijl}^{(w)} = K_{ij}) \right) \quad (6)$$

If state of any unit exceeds $L_{PM,ij}$, all the units in the subsystem whose state exceed $L_{OM,ij}$ and less than K_{ij} should be maintained preventively. So expectation of preventive cost of the subsystem can be expressed as

$$E_{CPM,ij}(\theta_{ij}) = C_{PM,ij} \sum_{w=1}^{S_{ij}} \left(\left(\pi_{ij}^{SR}(\theta_{ij}) \right)_w \times I(X_{ij,1}^{(w)} \geq L_{PM,ij}) \times \sum_{l=1}^{M_i} I(K_{ij} > X_{ijl}^{(w)} \geq L_{OM,ij}) \right) \quad (7)$$

Maintenance cost C_M of the system can be expressed as

$$C_M = \sum_{i=1}^N C_{M,i} \quad (8)$$

3.2 Availability of the system

N subsystems are connected in series, and UGF can be utilized to get

$$\begin{aligned} U(z) &= \Omega_{i=1}^N [U_{ij_i}(z)] = \Omega_{i=1}^N \left[\sum_{v_i=1}^{N_{PRO,ij_i}} A_{ij_i v_i}(\theta_{ij_i}) z^{W_{ij_i v_i}} \right] \\ &= \sum_{v_1=1}^{N_{PRO,1j_1}} \sum_{v_2=1}^{N_{PRO,2j_2}} \dots \sum_{v_N=1}^{N_{PRO,Nj_N}} \left[\left(\prod_{i=1}^N A_{ij_i v_i}(\theta_{ij_i}) \right) z^{\min\{W_{1j_1 v_1}, W_{2j_2 v_2}, \dots, W_{Nj_N v_N}\}} \right] \end{aligned} \quad (9)$$

$N_{PRO,ij}$ is the number of state for i th subsystem of version j . Eq. (12) can be expanded to be polynomial form and N_{PRO} is the number of state of the system, then

$$U(z) = \sum_{v=1}^{N_{PRO}} A_v(\theta) z^{W_v} \quad (10)$$

In literature (Nahas, et al., 2008), the whole running section was divided into p parts, each one of them proposes requirements of working efficiency W_{0p} and running time T_p . Availability of the system in single section could be defined as

$$A_p = \sum_{\{v | W_v \geq W_{0p}\}} A_v \quad (11)$$

Availability of the whole running section can be calculated by weighted average

$$A_s = \left(\sum_{p=1}^P A_p T_p \right) / \sum_{p=1}^P T_p \quad (12)$$

4. Hybrid algorithm of PSO and LS

4.1 Objective function

Eq. (13) is the model of a redundancy allocation problem (RAP) whose goal is to get the minimal installation and maintenance cost, meanwhile system availability A_s is higher than required value A_0 during the whole running interval.

$$\begin{cases} \text{Goal: } \min C_s = C_{IN} + C_M \\ \text{Constraint: } A_s \geq A_0 \end{cases} \quad (13)$$

One important problem in swarm optimization analysis is how to search in feasible region. Penalty function can be used to solve this problem effectively. So, objective function is redefined as the sum of itself and penalty function. Calculation method of penalty function is put forward below.

$$C_P = \begin{cases} 0 & , \text{If } A_S \geq A_0 \\ C_\gamma \tan^{1.5} [-0.5\pi(1 + A_S/A_0)], & \text{else} \end{cases} \quad (14)$$

Where C_γ is the penalty coefficient, after penalty is considered, objective function can be changed to

$$C_S = C_{IN} + C_M + C_P \quad (15)$$

4.2 Coding of RAP

According to the description of above model, the whole series-parallel system can be coded $[X, Y]$, X is the position vector and Y is the velocity vector. With regard to position vector $X = [X_1, X_2, \dots, X_N]$, Where $X_i = [j, M_i, U_{ij}]$, X_i represents the characteristic of the subsystem, j is the version, M_i is the amount of units, and U_{ij} is the index number that the inspection and maintenance strategy selected. Too many units will add the installation cost of the system and it's impossible to achieve the goal of low cost, so there is need to set an upper limit X_B to restrain the number of units in each subsystem.

4.3 Local search strategy

To compensate for the influence that discretization works on particle's position and velocity vector, local search method is applied on individual optimal solution X_{PB} and global optimal solution X_{GB} . To enhance calculation efficiency, local search is executed on X_{PB} that take up very little proportion η in every n iterations, while for X_{GB} , if there's a better one, local search will be executed on it.

Main idea of local search is that a solution with lower cost and higher availability. Research object of local search is characteristic of the subsystem which includes version, amount and index number of maintenance strategy. Version and amount of units should be enumerated firstly, maintenance strategy of the subsystem will be worked out by restricting maintenance cost, of which the upper limit is the minimum cost of the subsystem before this local search, when conducted local search at the first time, upper limit is the cost of characteristic vector originated from stochastic enumeration. Then Eq. (12) is used to calculate availability A_i of the subsystem, lastly, A_i will be ranked from high to low, then the top N_s particles and corresponding cost will be recorded. Characteristics of other subsystems stay unchanged when the i th subsystem undergoes local search. The particle with minimal cost is the optimal one to be output after all subsystems have been searched.

5. Analysis of examples

5.1 Typical problem

5.1.1 Problem description

Some auto parts factory plans to put a vehicle air condition valve plug production line into production which includes four main processes. Units in every process are uniform. There are four versions of units in every process. Working performance of units from each factory varies from each other. Performance degradation process of all units all follow discrete time Markov process. Discrete interval is 1 week. State transition probability matrix of units produced from these four factories in the first process can be expressed as

$$\begin{aligned} P_{11} &= \begin{bmatrix} 0.6 & 0.2 & 0.15 & 0.04 & 0.01 \\ 0 & 0.6 & 0.3 & 0.08 & 0.02 \\ 0 & 0 & 0.75 & 0.2 & 0.05 \\ 0 & 0 & 0 & 0.75 & 0.25 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} & P_{12} &= \begin{bmatrix} 0.6 & 0.2 & 0.15 & 0.03 & 0.015 & 0.005 \\ 0 & 0.6 & 0.3 & 0.06 & 0.025 & 0.015 \\ 0 & 0 & 0.75 & 0.2 & 0.03 & 0.02 \\ 0 & 0 & 0 & 0.75 & 0.15 & 0.1 \\ 0 & 0 & 0 & 0 & 0.7 & 0.3 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \\ P_{13} &= \begin{bmatrix} 0.7 & 0.15 & 0.1 & 0.04 & 0.01 \\ 0 & 0.65 & 0.25 & 0.08 & 0.02 \\ 0 & 0 & 0.6 & 0.25 & 0.15 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} & P_{14} &= \begin{bmatrix} 0.7 & 0.15 & 0.1 & 0.03 & 0.015 & 0.005 \\ 0 & 0.7 & 0.2 & 0.06 & 0.025 & 0.015 \\ 0 & 0 & 0.75 & 0.2 & 0.03 & 0.02 \\ 0 & 0 & 0 & 0.75 & 0.15 & 0.1 \\ 0 & 0 & 0 & 0 & 0.8 & 0.2 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

To reduce redundant description, state transition probability matrix of units in other processes are the same with that of the first process such as $P_{21}=P_{11}$, productivity and installation cost of each unit is listed in Table 1. Number of inspection threshold of units $N_{INS,ij}$ in each process is 2, then maintenance strategy includes $L_{PM,ij}$, $L_{OM,ij}$ and two inspection thresholds $L_{INS,d,ij}$, where $d=1,2$.

Table 1 Productivity, installation and maintenance cost of different versions in each subsystem

i	j	$\gamma_{ij}(1)$	$\gamma_{ij}(2)$	$\gamma_{ij}(3)$	$\gamma_{ij}(4)$	$\gamma_{ij}(5)$	$\gamma_{ij}(6)$	$C_{IN,ij}$	$C_{PR,ij}$	$C_{CR,ij}$	$C_{H,ij}$	$C_{INS,ij}$
1	1	0.2	0.15	0.1	0.05	0	-	1.56	0.686	0.312	0.156	0.052
	2	0.4	0.32	0.24	0.16	0.08	0	1.86	1.085	0.62	0.155	0.031
	3	0.24	0.2	0.16	0.08	0	-	2.16	0.756	0.36	0.216	0.036
	4	0.8	0.7	0.6	0.4	0.2	0	3.9	1.869	1.068	0.134	0.13
2	1	0.08	0.06	0.04	0.02	0	-	1.548	0.539	0.31	0.155	0.052
	2	0.25	0.2	0.15	0.1	0.05	0	2.748	1.603	0.916	0.229	0.046
	3	0.15	0.125	0.1	0.05	0	-	2.901	1.015	0.484	0.29	0.048
	4	0.6	0.525	0.45	0.3	0.15	0	4.101	4.55	1.64	0.205	0.21
3	1	0.24	0.18	0.12	0.06	0	-	0.642	0.364	0.128	0.064	0.021
	2	0.45	0.36	0.27	0.18	0.09	0	1.152	0.672	0.384	0.096	0.019
	3	0.54	0.45	0.36	0.18	0	-	1.602	0.56	0.267	0.16	0.027
	4	1.6	1.4	1.2	0.8	0.4	0	5.4	1.288	0.737	0.092	0.092
4	1	0.1	0.075	0.05	0.025	0	-	2.049	0.714	0.41	0.205	0.068
	2	0.125	0.1	0.075	0.005	0.025	0	1.935	1.127	0.645	0.161	0.032
	3	0.09	0.075	0.06	0.03	0	-	2.091	0.735	0.349	0.209	0.035
	4	0.56	0.49	0.42	0.28	0.14	0	3.57	3.92	1.428	0.179	0.18

Proportion and productivity requirements of busy season, normal season and slack season in one year are listed in Table 2. To set three different examples, availability of the system A_0 is set to be 0.9, 0.95 and 0.99 separately.

Table 2 Productivity requirements of the system

W_{op}	0.40	0.30	0.20
$T_p(\%)$	20	30	50

Some other parameters needed in optimal algorithm are listed in Table 3.

Table 3 Parameters used in optimization algorithm

X_B	$t_{\max}$	N_P	C_γ	η	n	N_S
6	50	500	2000	2%	10	5

5.1.2 Result analysis

The results deduced from optimization program are listed in Table 4, decode procedure of results is the inverse process of the coding problem referred in 3.2. Take that the availability is restricted to be 0.9 for example. Two units with the version of 3 are applied in the second process and the corresponding maintenance strategy is 24 [4, 3, 3, 2]. If the worst state $X_{ij1}^{(w)}$ of the units in the subsystem when they are tested is 3, inspection interval will be 1 week, if $X_{ij1}^{(w)}$ is 2, inspection interval will be 2 weeks, if $X_{ij1}^{(w)}$ is 1, inspection interval will be lasted to 3 weeks. When $X_{ij1}^{(w)}$ is no less than maintenance threshold 4, failure units should be replaced, as to other units in the same subsystem of which the state is no less than opportunistic maintenance threshold 3 should be maintained preventively. After maintenance, inspection interval should be judged according to state of the subsystem.

Table 4 Optimization results

NO.	A_0	A_s	$C_s(\$10,000)$	Optimal solution
1	0.9	0.9003	9.77330	[2,3,1,3,2,2,2,3,32,24,8,8]
2	0.95	0.9508	10.2374	[2,3,2,3,2,2,2,3,24,23,25,7]
3	0.99	0.9903	10.9847	[2,3,1,3,2,2,2,3,24,8,3,3]

5.2 Comparative example

Two examples are put forward in this paper to prove that the method in this paper is better. One explains that state-based inspection affects maintenance cost of the system. The other compares effects on maintenance strategy of the system from changeable structure and fixed structure. It is the occasion when maintenance strategy should be considered, either structure is selected before maintenance strategy or studied simultaneously.

5.2.1 Effect of state-based inspection

If $L_{INS,1,ij}=K_{ij}$ and $L_{INS,2,ij}=K_{ij}$, under the circumstances of no breakdown unit, inspection interval will be 3. While the units tested to be failure should be replaced, they will recover, inspection interval will also be 3. Therefore, inspection interval of all states is 3. Similarly, when $L_{INS,1,ij}=K_{ij}$ and $L_{INS,2,ij}=1$, inspection interval is 2, when $L_{INS,1,ij}=1$ and $L_{INS,2,ij}=1$, inspection interval is 1. Consequently, on the condition that the availability $A_0=0.9$, results of fixed inspection interval are compared with that of state-based inspection, which are shown in Table 5.

Table 5 Details of systematic cost under different maintenance strategy (\$10,000)

Inspection interval	Inspection cost	Halting cost	Corrective maintenance cost	Preventive maintenance cost	Installation cost	Total cost
3	1.4400	2.0682	0.5489	3.7058	8.40	16.1629
2	1.3850	1.3449	0.3818	2.2997	7.86	13.2714
1	2.5600	1.2502	0.2104	1.2586	5.08	10.3592
State-based	2.0797	1.0857	0.2800	1.2179	5.11	9.77330

5.2.2 Structure-based effects

Fixed system structure is as shown in Fig.1, versions applied are [3 2 4 2] separately. After optimization, results under different availability are shown in Table 6.

Table 6 Optimization results of fixed structure

Required availability	Actual availability	Cost(\$10,000)	Maintenance strategy
0.90	0.9024	13.4018	8 25 15 27
0.95	0.9506	13.9858	10 25 25 13
0.99	0.9907	15.0865	10 24 10 10

5.2.3 Results analysis

Details of the cost used in the system are listed in Table 5. All the cost is the mean value in unit time. Preventive maintenance cost includes preventive and opportunistic maintenance cost. As is shown in these two tables, decline of the inspection interval infers that ability or amount of the units in the same process descends properly. Although reducing the inspection interval will enhance inspection cost, it is far less than the cost of other items. Compared with fixed inspection interval, state-based inspection maintenance strategy is more flexible in determining inspection time and reducing running cost of the system. Results of maintenance cost of the fixed system structure with different availability are listed in Table 6. The method in this paper takes the influence that maintenance strategy acts on the system into consideration before determination of the system structure. The combination of the minimum running cost will be found on this basis. Compared with traditional method that maintenance strategy is optimized on the base of known system structure, our method highlights the effect that maintenance acts on system structure and then even more optimal system configuration will be found.

6. Conclusions

A novel series-parallel system design method is put forward in this paper under the consideration of system structure, preventive maintenance and state-based inspection. Formulations of installation and maintenance cost as well as availability of the system are deduced through Markov process combined with UGF method. A hybrid algorithm is put forward on the basis of PSO and LS, which is mainly used in integer domain compared with standard PSO method. LS method can search more carefully around individual and global optimal solution. Finally, results of the example show that:

(1) Appropriate inspection will be helpful in controlling state of the system and reducing maintenance cost caused by improper maintenance.

(2) Combining system structure with maintenance strategy will be obviously more efficient than separate consideration.

Acknowledgement

This study was supported by the Fundamental Research Funds for the Central Universities of China (Grant No. N140303010).

References

- Abdelbar AM., & Abdelshahid S. (2004) Instinct-based PSO with local search applied to satisfiability. *2004 Ieee International Joint Conference on Neural Networks, Vols 1-4, Proceedings*, 2291-2295.
- Clerc M & Kennedy J. (2002) The particle swarm - Explosion, stability, and convergence in a multidimensional complex space. *Ieee Transactions on Evolutionary Computation*, 6: 58-73.
- Kuo W, & Wan R (2007). Recent advances in optimal reliability allocation. *Ieee Transactions on Systems Man and Cybernetics Part a-Systems and Humans*, 37: 143-156.
- Nahas N, Khatab A, Ait-Kadi D, & Nourelfath M (2008). Extended great deluge algorithm for the imperfect preventive maintenance optimization of multi-state systems. *Reliability Engineering & System Safety*, 93: 1658-1672.
- Park JB, Lee KS., Shin JR., & Lee KY (2005). A particle swarm optimization for economic dispatch with nonsmooth cost functions. *Ieee Transactions on Power Systems*, 20: 34-42.
- Shafiee M, & Chukova S (2013). Maintenance models in warranty: A literature review. *European Journal of Operational Research*, 229: 561-572.
- Kuo W, & Zuo MJ. (2003). Optimal Reliability Modeling: Principles and Applications. Wesley.
- Wang Y, Li L, Huang SH, & Chang Q (2012). Reliability and covariance estimation of weighted k-out-of-n multi-state systems. *European Journal of Operational Research*, 221: 138-147.
- Yan SY, Shih YL, & Lee WT (2011). A particle swarm optimization-based hybrid algorithm for minimum concave cost network flow problems. *Journal of Global Optimization*, 49: 539-559.
- Zhou YF, & Zhang ZS (2013). Maintenance optimisation of a multi-state series-parallel system considering economic dependence and state-dependent inspection intervals. *Reliability Engineering & System Safety*, 111: 248-259.

A New Modelling Methodology for the Assessment of Maintenance Policies of Subsea Systems

Y. Zhang*, A. Rauzy **, A. Barros***,

IPK, Norwegian University of Science and Technology, S.P. Andersens veg 5, 7491 Trondheim, Norway.

*Email address: *yun.zhang@ntnu.no, **antoine.rauzy@ntnu.no, ***anne.barros@ntnu.no*

Abstract In this article, we propose a new modelling methodology for the assessment (and later on optimization) of inspection and maintenance policies of oil and gas subsea systems. This new methodology relies on the high level modelling language AltaRica 3.0. It makes it possible to model accurately the specific features of inspection and maintenance policies of subsea systems while facilitating knowledge capitalization via the reuse of modelling components and modelling patterns.

1. Introduction

Subsea systems are prone to degradations and failures because they are located in harsh environment. Therefore, they are designed essentially not to be maintained (Moreno-Trejo and Markeset (2011)). Still periodic inspections and maintenances (mainly retrievals of modules) are performed. These interventions are subject to severe constraints. In practice, they are the main responsible for systems' downtime. As of today, inspection dates are essentially scheduled at design phase, thus obeying a very conservative approach. It would be of interest to put in place a more dynamic approach based on monitoring of components and observations made during inspections. Such a revision of current practices needs to be strongly supported by evidences, i.e. in a first phase by off-line analyses.

Classical modelling formalisms used in RAMS analyses do not suffice for that purpose: combinatorial formalisms such as Fault Trees or Reliability Block Diagrams are not expressive enough to represent complex maintenance policies applied in subsea systems. (Multiphase) Markov chains are unmanageable (by hand) when the system under study is made of more than a few components. Finally, stochastic Petri net model tends to be hard to design and even harder to maintain and to share amongst stakeholders (see e.g. Signoret et al. (2013) for a discussion on this topics).

The high level language AltaRica 3.0 (see e.g. Prosvirnova (2014)) does not suffer from these limitations. Its underlying mathematical framework, guarded transition systems (Rauzy (2008)), makes it possible to represent both states of components and transitions amongst states under the occurrence of events and propagation of information through the network of components. With that respect, guarded transition systems generalize stochastic Petri nets (at no algorithmic cost). Moreover, AltaRica 3.0 model structuring constructs, stemmed from prototype-oriented programming (Noble et al. (1999)), makes it possible to design libraries of reusable modelling components and patterns, thus facilitating knowledge capitalization.

This said, subsea systems have very specific characteristics and we have to define a modelling strategy for our case study, namely a subsea high pressure protection system. The behaviour of basic components such as pressure transmitters, valves and CPU can be represented classically by means of finite state automata. Inspection and maintenance policies have to be described by a controller of some sort, that can also be encoded by means of a finite state automaton. But the question is how to relate automata for the basic components with the automaton of the controller? We propose here to describe this relation by means of three successive flows of information: basic components inform the controller of their degradation level, the controller orders basic components to switch from different modes (operation, test, maintenance), and finally basic components export their actual production so to describe the availability of the system as a whole. The interesting point here is that these flows propagate through different hierarchical decompositions of the system: because basic components are not maintained individually but by means of retrievals of modules, the hierarchical decomposition to synthesize degradation indicators is not the same as the hierarchical decomposition to determine the availability of the system.

Once the system is fully described using the above modelling methodology, the mean availability and down-time of the system can be assessed by Monte-Carlo simulation.

The contribution of this article is thus twofold. First, it shows how AltaRica 3.0 can be advantageously used to model maintenance policies for non-trivial (subsea) systems. Second, it proposes a new modelling methodology that helps to design and to reuse maintenance models. This paves the way condition monitoring and optimization of maintenance policies for a large class of systems.

The remainder of this article is organized as follows. Section 2 presents the case study. Section 3 describes the proposed modelling methodology. Section 4 discusses its implementation in AltaRica 3.0. Finally, Section 5 concludes the article and gives some perspectives.

2. Case study

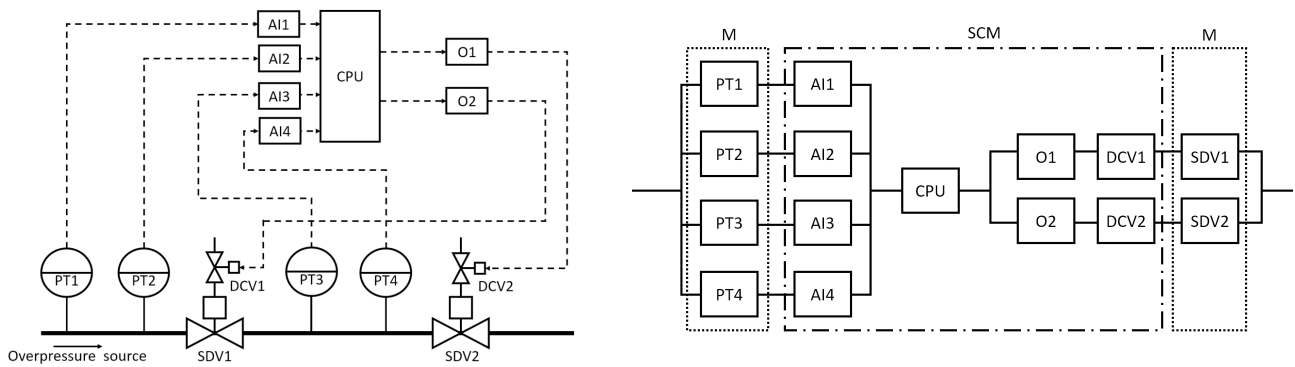


Figure 1. Left: a typical HIPPS in subsea system; Right: reliability block diagram (RBD) for the HIPPS function.

Figure 1 (Left) shows a (simplified) high integrity pressure protection system (HIPPS) of a subsea oil and gas production plant. HIPPS is a safety instrumented system (SIS) that isolates high-pressure source from low-pressure flowlines in the occurrence of a high-pressure event (e.g. loss of control of wells, inadvertent valves closure, blockage etc.). The above HIPPS consists of 15 components that can be split into three main categories: sensors, logic solver and final elements. The four pressure transmitters (PTs) are connected to a CPU via respective analogue inputs (AIs). The CPU decides based on 2-out-of-4 (2oo4) voting rule whether or not to activate the final element via two redundant outputs (Os). The final elements are composed of two directional control valves (DCVs) and two shutdown valves (SDVs) that stop the flow from upstream source (left) to downstream facilities (right) thus avoiding a loss of containment in overpressure situations. The shutdown valves act at the same time even only one is necessary to stop the flow. The CPU, AIs, Os and DCVs are located in the subsea control module (SCM), while the PTs and SDVs are placed on the pipelines on the manifold (M), as shown in Figure 1 (Right).

Given the system description and inputs from API-RP-170 (2014), IOGP-RP-2015 (2015) and Bak et al. (2007), we can make the following assumptions.

All components are subject to failures. As a first approximation, the failure of the component C ($C \in \{PT, AI, CPU, O, DCV, SDV\}$) is assumed to be exponentially distributed, with a rate λ_C . This failure rate includes the effects of monitoring, partial testing and the like.

It is not possible to repair components individually. Preventive and corrective maintenance actions require to retrieve the whole module the components belong to, i.e. either to subsea control module or the full manifold, and to replace it. After a replacement, modules can be considered as good as new. Both retrievals require to shut down the production. Retrieving the subsea control module and the manifold takes respectively times ρ_{SCM} and ρ_M . Obviously, ρ_M is much bigger than ρ_{SCM} .

Components CPU, AIs, Os and DCVs are continuously monitored. SDVs are periodically tested with a full functional testing. In practice, the period π of these tests is about one year. The test is assumed to fully cover potential problems. It requires to shut down the production for a time τ . PTs are tested at the same time as

SDVs, but sensor lines (i.e. pairs PT/AI) are continuously checked for consistency. If one of these lines is detected as failed, the system is reconfigured into a 1-out-of-3 logic.

There is a quite large delay δ between a maintenance operation is decided and it is actually started (typically about one month). This is the time required to mobilize the maintenance team, the vessel (including the remote operated vehicle) and replacement modules.

The states of the system can be split into three categories: the working state in which the system can be operated without any restriction, the failed or nearly failed state in which operations must be stopped, and degraded state in which a maintenance operation must be scheduled but the system can still be operated, possibly with some restrictions and/or with a decreased production.

For a given technology and in given exploitation conditions, the parameters on which the operator can actually act are the inspection period π and the characterization of the “degraded” category. The objective is to minimize downtimes due to inspection and maintenance, while preserving a sufficient availability on demand of the SIS (which is classified SIL4 in our case).

3. Modelling methodology

As we can see from Figure 1 (Right), the difficulty of modelling the HIPPS lies in the difference between its functional structure and maintenance decision process. As an illustration, consider the sensing line made of PT1 and AI1. Both components are required to work for this line to work. However, they belong to different modules. Therefore, their replacements do not involve the same operations. The availability of the system depends on the performance of each components according to the functional RBD structure in Figure 1 (Right). The performance of components depends itself on decisions made according to their degradation level. But these decisions are made considering the system as a whole, via the modules SCM and M delimited by dashed lines on Figure 1 (Right). Therefore, two different hierarchical decompositions of the system are necessary to represent the HIPPS: a decomposition to describe degradation levels and control and another decomposition to describe the availability of the system. Both of the decompositions are built over the same set of basic components, but aggregate and process different information flows for their own purposes (Figure 2).

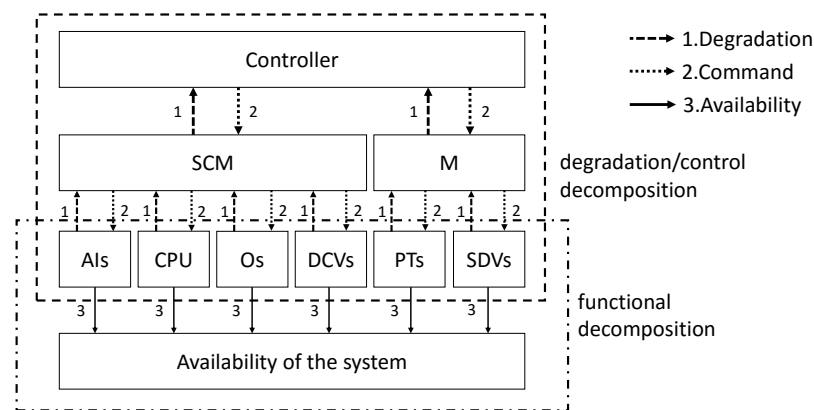


Figure2. A view of two hierarchical decompositions of the system.

The core of our model is a automatic controller which decides “when to do what”. The internal mechanism of the controller is shown in Figure 3.

The controller imports a condition information from lower level components (in the degradation/control hierarchy), i.e. SCM and M. Condition information on the SCM is updated continuously by monitoring components. The condition information on M is updated periodically by tests. Here we assumed that the condition information is one of the three values “WORKING”, “DEGRADED” or “FAILED”. Note that this information could be described as well as a degradation level (i.e. an integer or a real). To keep the figure simple, we assumed also that the controller receives this information already synthesized from modules SCM and M.

As discussed in section 2, the HIPPS can be mainly in three states “WORKING”, “DEGRADED” or “FAILED”) and three modes (“OPERATION”, “TEST” and “MAINTENANCE”). Therefore, the system may

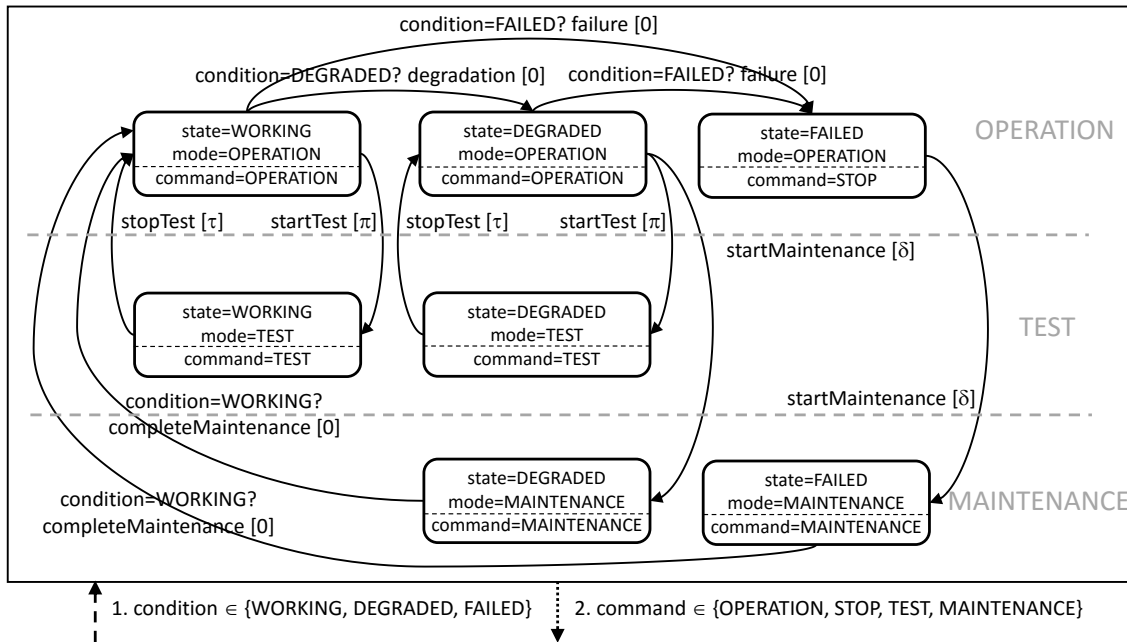


Figure3. Controller for the HIPPS.

be in one of the seven conditions (combinations of a state and a mode) represented by rounded rectangles on the figure.

In each of these conditions, the controller elaborates a command (indicated under the dashed line on figure). In modes “TEST” and “MAINTENANCE”, the command is just the mode. In mode “OPERATION”, the command depends on the state because the whole production has to be shut down when the HIPPS is failed.

Transitions between conditions are represented by arrows and triggering events are shown in text next to each arrow. Some of these transitions are conditioned by the condition information the controller receives. The delays associated with transitions are indicated within square brackets.

The system can only degrade and fail during operation. When the controller is in operation mode, it changes of state instantaneously when it receives information that the system is degraded or failed. It switches to the mode “TEST” every π time units. So if the controller enters the “OPERATION” mode at time t , it switches to “TEST” mode at time $t + \pi$. The test duration (i.e. the delay before the controller goes back to operation mode) is τ . Tests are performed only when the system in state “WORKING” or in state “DEGRADED”.

Once the controller detects that the system condition enters a degraded but workable level or into a failure state, it launches the command for a maintenance operation. There is a deterministic delay δ before the maintenance actually starts. If the system is only degraded, the production can go on, possibly at a decreased rate, otherwise it has to be stopped (hence the two different commands).

The system is shut down during the maintenance mode and fully recovered its “WORKING” state after maintenance. The controller goes back to “OPERATION” mode instantaneously. Of course, retrieving modules takes time, but it is assumed here that this time is defined at module level.

All the delays associated with the transitions of the controller are deterministic. Delays associated with transitions of modules and components are rather stochastic, because these transitions represent failures and maintenance actions.

4. Implementation in AltaRica 3.0

Two fundamental mechanisms are needed to implement the model proposed in the previous section: first, a mechanism to describe states and transitions, as for instance in Markov chains and Petri nets; second, a mechanism to propagate information in the network of components, as in Fault Trees or Reliability Block Diagrams.

Guarded Transition System (GTS) (Rauzy (2008)), which are at the core of the high modelling language AltaRica, provides these two mechanisms.

Formally, a GTS is a quintuple $\langle V, E, T, A, \iota \rangle$ where:

- V is a finite set of variables. V is the disjoint union of two sets: the set S of state variables and the set F of flow variables. Variables take their values into sets of constants, called them domain. Domains can be Boolean, integers, real numbers, sets of symbolic constants etc.
- E is a finite set of events that may occur in the system. Events are associated with delays that can be deterministic or stochastic.
- T is a set of transitions $G \xrightarrow{e} P$, where e is an event in E , G is a Boolean condition over state and flow variables called the guard of the transition, and P is an instruction that modifies the values of (some) state variables called the action of the transition.
- A is an assertion, i.e. an instruction that calculates the values of flow variables from the values of state variables.
- ι is the initial value of variables.

The current state of the system is described by the value σ of variables. A transition $G \xrightarrow{e} P$ is fireable in the current state when its guard G is satisfied by σ , i.e. $G(\sigma) = \text{true}$. Firing the transition consists in two steps: first, the action P is applied to state variable, then the assertion A is applied to update the value of the flow variable. The resulting state is thus $A(P(\sigma))$. A is actually a fix point calculation, making it possible to propagate changes of the values of state variables through the network of components, even in presence of loops (which is not the case in our model however).

As an illustration, consider the controller pictured in Figure 3. The AltaRica code for this controller is given as Figure 4. This code starts by three domain declarations. Then, it declares a block, i.e. a component, representing the controller. Each block encodes a GTS.

AltaRica makes it possible to compose GTS and to declare blocks as classes that can be then instantiated in a model. An AltaRica model is thus a hierarchy of blocks (and instances of classes). For our case study, the model involves a block for each basic components, for the modules SCM and M, and for the controller. These blocks are connected by means of flow variables (as the variables `condition` and `command` in the block of Figure 4) and assertions.

Once the full model assembled, various probabilistic indicators of interest, such as the probability of failure or the mean downtime of the system, can be assessed by means of Monte-Carlo simulations.

5. Conclusion

In this article, we proposed a new modelling methodology for the assessment (and later on optimization) of maintenance policies of oil and gas subsea systems. The model is a hierarchy of finite-state automata linked with information flows. Each automaton can be reusable and assembled into a larger or more complex system depends on users' needs. What the users need to consider only is which flow of each automaton is connected to those of the others. The fundamental idea here is to organize the model around three flows propagating through the network of components: a flow to synthesize degradation levels, a flow to send commands and finally a flow to describe the availability of the system. The description of these flows required different hierarchical decompositions of the system.

Acknowledgement

This work, which is a part of the PhD work of the first author, is supported by the SUBPRO project funded by the Norwegian Research Council.

References

- Jorge Moreno-Trejo and Tore Markeset. Identifying challenges in the maintenance of subsea petroleum production systems. In *Advances in Production Management Systems. Value Networks: Innovation, Technologies, and Management*, volume 384 of *IFIP Advances in Information and Communication Technology*, pages 251–259. Springer-Verlag, Berlin and Heidelberg, Germany, 2011. ISBN 978-3-642-33979-0. doi: 10.1007/978-3-642-33980-6_29.
- Jean-Pierre Signoret, Yves Dutuit, Jean-Pierre Cacheux, Cyrille Folleau, Stéphane Collas, and Philippe

```

domain Condition {WORKING, DEGRADED, FAILED}
domain Command {OPERATION, STOP, TEST, MAINTENANCE}
domain Mode {OPERATION, TEST, MAINTENANCE}

block Controller
  Condition state (init = WORKING);
  Mode mode (init = OPERATION);
  Condition condition (reset = WORKING);
  Command command (reset = OPERATION);
  event degradation (delay = 0);
  event failure (delay = 0);
  event startTest (delay = pi);
  event stopTest (delay = tau);
  event startMaintenance (delay = delta);
  event completeMaintenance (delay = 0);
  parameter Real pi = 8760;
  parameter Real tau = 12;
  parameter Real delta = 720;
  transition
    degradation: state==WORKING and mode==OPERATION and condition==DEGRADED
      -> state:=DEGRADED;
    failure: state!=FAILED and mode==OPERATION and condition==FAILED
      -> state:=FAILED;
    startTest: state!=FAILED and mode==OPERATION -> mode:=TEST;
    stopTest: mode==TEST -> mode:=OPERATION;
    startMaintenance: state!=WORKING and mode==OPERATION -> mode:= MAINTENANCE;
    completeMaintenance: mode==MAINTENANCE and condition==WORKING
      -> {state:=WORKING; mode:=OPERATION; }
  assertion
    command := if mode==OPERATION and state==FAILED then STOP else mode;
end

```

Figure4. AltaRica 3.0 code for the Controller.

- Thomas. Make your petri nets understandable: Reliability block diagrams driven petri nets. *Reliability Engineering and System Safety*, 113:61–75, 2013. doi: doi:10.1016/j.res.2012.12.008.
- Tatiana Prosvirnova. *AltaRica 3.0: a Model-Based Approach for Safety Analyses*. Thèse de doctorat, École Polytechnique, Palaiseau, France, November 2014.
- Antoine Rauzy. Guarded transition systems: a new states/events formalism for reliability studies. *Journal of Risk and Reliability*, 222(4):495–505, 2008. doi: 10.1243/1748006XJRR177.
- James Noble, Antero Taivalsaari, and Ivan Moore. *Prototype-Based Programming: Concepts, Languages and Applications*. Springer-Verlag, Berlin and Heidelberg, Germany, 1999. ISBN 978-9814021258.
- API-RP-170. *API Recommended Practice 170: Recommended Practice for Subsea High Integrity Pressure Protection System (HIPPS), Second Edition*. American Petroleum Institute, 2014. distributed by Thomson Reuters.
- IOGP-RP-2015. High integrity protection systems - recommended practice. Technical Report 443, International Association of Oil & Gas Producers, February 2015.
- Lars Bak, Roald Sirevaag, and Halvor Stokke. Hipps protects subsea production in hp/ht conditions. *Offshore, online magazine*, 67(6), 2007.

Analysis of Risk Events in UAV Flight Test Project based on Bayesian Network

Su Zhenfei *, Liu Baocheng ** and Ma Lin ***

* 37 Xueyuan Road, Haidian District, Beijing 100191, China + email address: szfei27@outlook.com

** Aviation Industrial Park, Jinwan District, Zhuhai City, Guangdong Province, 519040, China + email address: liubaocheng690113@163.com

*** 37 Xueyuan Road, Haidian District, Beijing 100191, China + email address: malin@buaa.edu.cn

Abstract. Unmanned aerial vehicle (UAV) is playing an increasingly important role in both military and civilian field. However, there could be great risks in the scientific research projects for UAV because of the features of the aviation products, especially in the flight test phase. Therefore, studying on the risk of UAV flight test projects is of great significance. The method proposed in the paper comprises three steps: identifying the risk events based on operational process analysis, building the Causal Bayesian Network (CBN) model of the risk events identified in the previous step, and determining the risk level of risk events and the key risk events by analyzing the built CBN model. The method focuses on analyzing causality among risk events. Bayesian Truth Serum (BTS) is used to determine the parameters of CBN, which can eliminate the deviation in expert judgment effectively. The paper uses Agenarisk to build and analyze the CBN model and the method is applied to and verified in the flight test of a certain type of UAV. The method helps to analyze the causality among risk events and to determine the key risk events, and then provides the basis for making effective risk response plan.

1. Introduction

Military UAV is becoming more and more important in the weapon equipment system because of its advantages of good concealment, no casualties and flexible operation. UAV is also playing an increasingly important role in civilian field such as in meteorology and agriculture.

UAV flight test is a key step throughout the process of development of UAV technology, including the research, designing, development, production and service of UAV. Problems could be found through flight test, which would help to improve the performance and the reliability of UAV.

It should be noticed that there could be great risks in the scientific research projects for UAV because of the features of the aviation products, especially in the flight test phase. Problems occurred in the flight test phase may lead to equipment damage which could result in considerable economic losses. Therefore, studying on the risk of UAV flight test projects is of great significance.

In the risk management of UAV flight test, common methods such as risk matrix and probabilistic risk analysis have the roughly same thought, based on the process "to identify the possibilities of risk events - estimating probability distribution and affect of random events – solving the affect of total risk event". Recent study (Hubbard DW, 2009) (Cox LA, 2011) shows that these methods look all risk events as isolated random events, and often ignore the mutual influence among the risk events, including causal relationship. Vose. D. (2008) have stated that some risk management methods analyze the correlation among risk events to some extent, but the statistical correlation among risk events based on statistical methods can not take the place of the causal relationship among risk events.

In addition, the evaluation the possibility of risk events and its consequences is often highly dependent on expert judgment. A large number of studies (Edwards W *et al*, 2007) (Kynn M, 2008) have shown that there may a lot of deviation in expert judgment and most of the risk analysis methods ignore processing for the deviation.

In this paper, CBN is used for modeling of risk events and the causal relationship among risk events of UAV flight test is identified effectively. BTS is used to estimate the parameters, which eliminates the possible deviation in expert judgment.

2. Analysis of Risk Events based on Casual Bayesian network

We look causality as a sufficient condition, which means that happening of the reason event A must lead to the result event B. And the reason event A must happen before the result event B. In the field of risk

management, most causality is embodied in a probabilistic causal relationship, meaning that there's a high probability of a particular result event if the reason event happens. This knowing avoids many debates about causality which still have no conclusion today, and also has important practical value of achieving the goal of risk management.

Causal analysis is to determine the cause of a phenomenon changing, mainly to solve the problem of "why". Causal analysis studies events which happen before the target event and distinguishes the reason events from the non-reason events. Or it studies events which happen after the target event and distinguishes the result events from the non-result events. In risk management, we can take advantage of causal analysis to do risk diagnosis. When risk event A happens, we can find out the main possible reasons for event A. Or we can use causal analysis to do risk prediction, that is, if risk event A happens, we can deduce how event A may affect other events. Therefore, understanding the causal relationship among risk events is very necessary.

Causal Bayesian Network (CBN) is a kind of directed acyclic graph, whose nodes and directed edges can be used to represent risk events and causal relationship among risk events separately. Pearl J. (2009) have proposed that CBN is one of the most effective models used to analyze causality. We adopt CBN to describe and evaluate the causal relationship among risk events in the paper.

Analysis of risk events based on CBN includes three steps, the first is to identify risk events by operational process analysis, and then to construct CBN model and finally to analyze the model to determine the risk level of total risk event and key risk events.

2.1 Identifying risk events based on operational process analysis

Identifying risk events based on operational process analysis on the one hand emphasizes occurring sequence and occurring stages of risk events, advantageous to analyze causal relationship among risk events and to determine the structure of CBN, on the other hand, provides basis for subsequent risk management to establish risk responsibility.

2.2 Causal Bayesian Network modeling

In CBN, causal relationship can be transformed to conditional independence assumption, and conditional probability tables can be transformed to causal contribution of parent nodes (reasons) to child nodes (results), which simplifies the building and reasoning of CBN greatly.

CBN modeling is made up of two parts, determination of the structure and parameters, corresponding to the qualitative description and the quantitative description.

In the paper, risk events are identified through operational process analysis, and the structure of CBN can be built based on this step.

CBN has two parameters, namely nodes parameters and directed edges parameters. When using CBN for risk management, node parameters represent the risk level of risk events. Risk level is a comprehensive qualitative measurement of probability and consequences of the risk events. In the paper risk level is divided into five levels, very high, high, medium, low and very low. Directed edge parameters represent causal contribution of parent-node risk events to child-node risk events, meaning once the parent-node risk events happen, how much impact will bring to the child-node risk events. Causal contribution is divided into five levels according to the effect degree, very high, high, medium, low and very low, respectively represented by integers from 1 to 5.

There are many methods used to estimate the parameters of CBN. Expert judgment is usually adopted due to the lack of data or high costs of data acquisition in the field of risk management, but the existing research (Edwards W *et al*, 2007) (Kynn M, 2008) shows that there may a lot of deviation in expert judgment.

Prelec D. (2004) put forward Bayesian Truth Serum (BTS) to eliminate the deviation in expert judgment. BTS not only asks respondents to choose the option which they think is right, also to estimate the probability of each option being selected by all respondents. Estimation of other respondents' choices is used to correct expert judgment. BTS is more scientific compared with other methods that choose the majority opinion as a result.

The paper uses BTS to judge the risk level. The number of experts is n , the number of options for risk level is five, represented by m . Define $x_k^r \in \{0, 1\}$ as whether expert- r choose level- k , and define $y = (y_1^r, \dots, y_m^r)$ as expert- r 's prediction of the probability of a certain risk event on each risk level from all expert judgment. It can be known that $(y_k^r \geq 0, \sum_{k=1}^m y_k^r = 1)$ from the definition of y_k^r .

BTS is composed of four steps :

(1) Calculate the average of real choice $\bar{x}_k$ and the geometric average of the prediction $\ln \bar{y}_k$.

$$\bar{x}_k = \frac{1}{n} \sum_{r=1}^n x_k^r \quad (1)$$

$$\ln \bar{y}_k = \frac{1}{n} \sum_{r=1}^n \ln y_k^r \quad (2)$$

(2) Calculate the BTS score of every expert.

$$u^r = \sum_{k=1}^m x_k^r \ln \frac{\bar{x}_k}{y_k^r} + \sum_{k=1}^m \bar{x}_k \ln \frac{y_k^r}{\bar{x}_k} \quad (3)$$

(3) Calculate the BTS score of every risk level.

$$\bar{u}_k = \frac{1}{n \bar{x}_k} \sum_{r=1}^n x_k^r u^r \quad (4)$$

(4) Choose the risk level which maximizes $\bar{u}_k$.

All risk level of the last-level risk events and all causal contribution parameters among risk events are determined by expert judgment. Then the risk level of penultimate -level of risk events are determined based on risk level of last-level risk events and all causal contribution parameters obtained in previous step. And so on the risk level of the total risk event is obtained in this way.

2.3 Determining the risk level of total risk event and the key risk events

AgenaRisk is used to calculate the probability of the total risk event on each risk level. We can also determine the key risk events by sensitivity analysis using Agenarisk. In actual risk management, it could cause very high costs to control all risk events due to the number of risk events is so large. So the identification of key risk events appears especially important.

3. Analysis of risk events in UAV flight test project

The paper analyzes the risk events of type-A UAV flight test as an example. It's known that type-A UAV have completed early product development and test of some scientific research items. Type-A UAV flight test is to be carried out to verify the design and to improve the performance.

3.1 Identification of risk events in type-A UAV flight test

When analyzing the operational process of type-A UAV flight test, we should determine the milestones firstly and then determine specific assignments gradually, from top to bottom. It's known that type-A UAV flight test mainly includes three stages, stage of preparation, stage of flight test and stage of data processing. The operational process of type-A flight test is shown in figure 1.

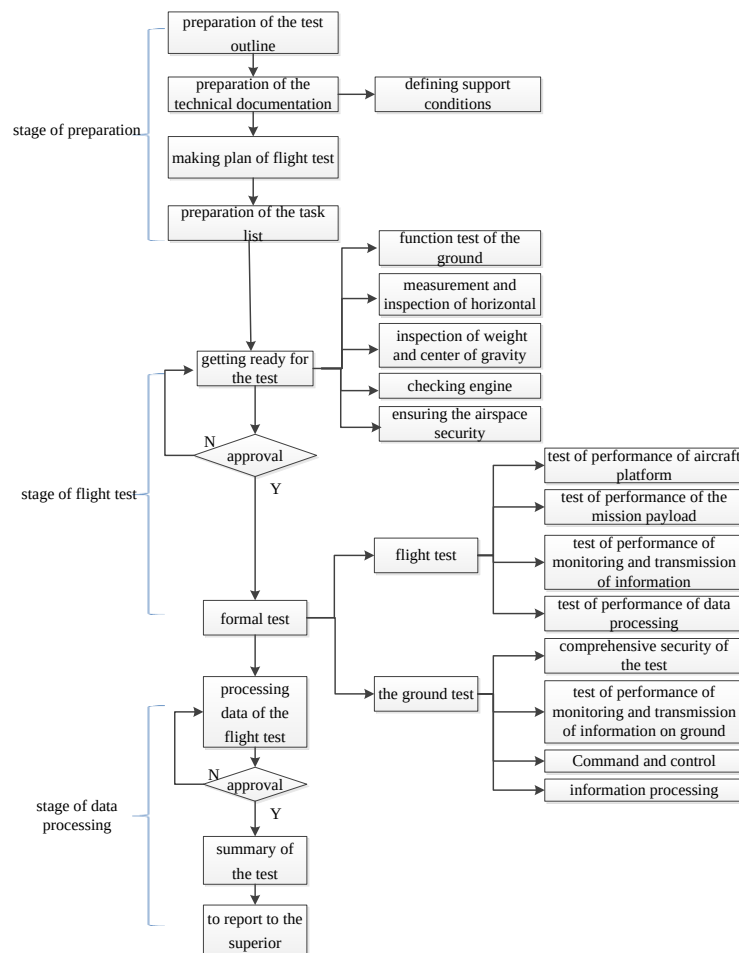


Figure 1. Operational process of type-A UAV flight test

Possible risk events identified by analyzing the operational process are shown in table1.

Table 1. Risk events in type-A UAV flight test

Fist-level risk events	Second-level risk events	Codename
Risk in preparing the test outline	Purpose or subjects of the test are not clear	A1
	Technical requirements are not reasonable	A2
	Assessment standard are Invalid	A3
Risk in preparing the technical documentation	Restrictions for outside are too high or too low	B1
	Technical conditions are not clear	B2
	Guarantee conditions are not clear	B3
Risk in plan and task list	Personnel responsibilities are not clear	C1
	The task and purpose are not consistent	C2
Risk in prior work	Prior test are not conducted according to the requirements	D1
	Failures are not returned to zero	D2
Risk in formal flight test	System failure	E1
	Human error	E2
	The airspace conditions are not appropriate	E4
Risk in processing data	Errors appear when processing data	F

3.2 CBN modeling of risk events in type-A UAV flight test

The structure of CBN transformed from risk events is shown in figure 2. Nodes in the figure are the total risk event, first-level risk events and second-level risk events of type-A UAV flight test.

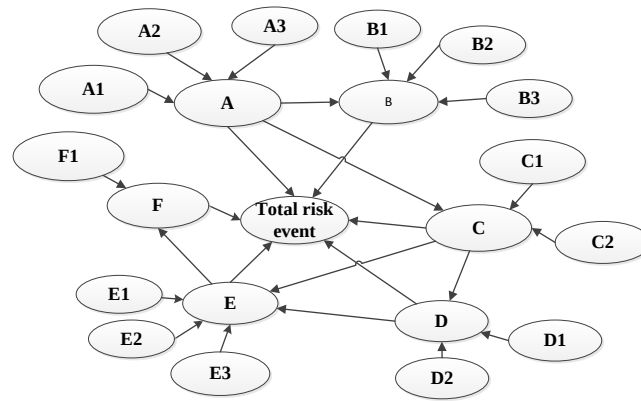


Figure 2. The structure of CBN of type-A UAV flight test

We select 15 experts to estimate risk level of the risk events, and use BTS to deal with the data. The risk level of second-level risk events is shown in table 2.

Table 2. Risk level of second-level risk events

	Risk level of second-level risk events		
A	A1	A2	A3
	very low	very low	low
B	B1	B2	B3
	low	very low	low
C	C1	C2	
	medium	very low	
D	D1	D2	
	low	low	
E	E1	E2	E3
	very high	low	medium
F	F1		
	low		

Causal contribution parameters among risk events are estimated by 15 experts, and average of their estimation is taken as the results of expert group judgment. The results are shown in table 3.

Table 3. Causal contribution parameters among risk events

Child-node risk events	Parent-node risk events					
A	A1	A2			A3	
	2.9	2.5			2.9	
B	B1	B2	B3		A	
	3.8	3.6	3.6		3.2	
C	C1		C2		A	
	4		1.6		2.6	
D	D1		D2		C	
	4		3.6		2.6	
E	E1	E2	E3	C		D
	4.4	4.4	4.4	2.6		4
F	F1			E		
	3.2			1.8		
Total risk event	A	B	C	D	E	F
	4.2	4.2	3.6	4.6	5	3.8

3.3 Risk level of total risk event and key risk events in type-A UAV flight test

Input the structure and parameters of CBN of type-A UAV flight test into Agenarisk, and the result is shown in figure 3. According to figure3, the probability of the total risk event being on low risk level is 94.531%. There's a very low probability when the total risk event is on other risk level.

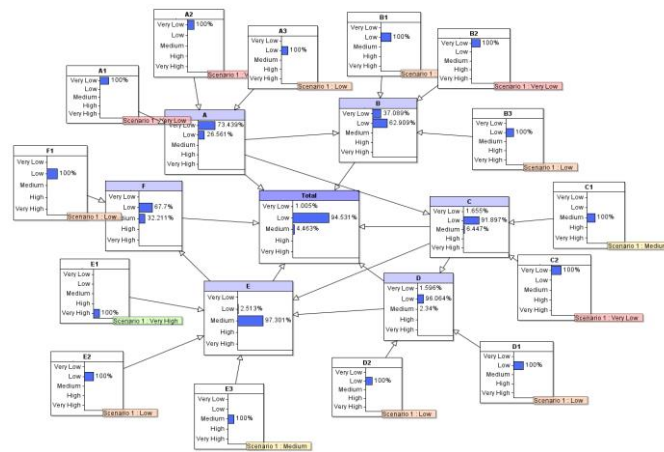


Figure 3. The risk level of type-A UAV flight test

Carry out sensitivity analysis on Agenarisk considering the condition of the risk level of the total risk event being low. Set the total risk event as the target node, and A1, A2, A3, B1, B2, B3, C1, C2, D1, D2, E1, E2, E3 and F1 as sensitivity analysis nodes. It can be concluded from figure 4 that C1 impacts the total risk event most, followed by A1, A3, F1. The control of these risk events should be especially strengthened.

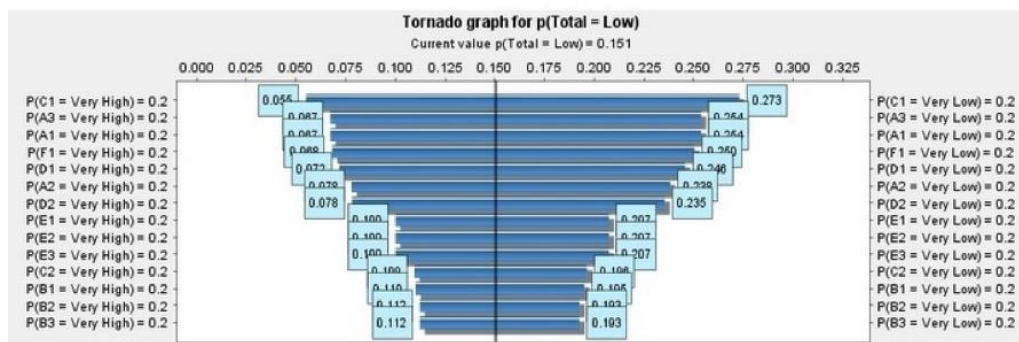


Figure 4. Sensitivity analysis of type-A UAV flight test

4. Conclusions

The analysis method of risk events put forward in the paper focus on causality among risk events. Risk identification based on operational process analysis corresponds risk events to operational process, advantageous to the analysis of the causal relationship among risk events. Modeling of risk events based on CBN can reflect the causality among risk events effectively. In addition, the use of BTS eliminates the possible deviation in expert judgment.

The paper takes type-A UAV flight test as a case study, have determined the risk level of the total risk event and the key risk events of type-A UAV flight test. First-level risk events and second-level risk events are considered in the case study. And the method of analyzing risk events put in the paper also can be applied to the case of more complex dependences among risk events, it focuses on causality among risk events and makes every step of the process verified, helping to achieve continuous improvement of risk management.

Reference

- Hubbard DW. (2009) The Failure of Risk Management: Why It's Broken and How to Fix It. *Risk Analysis* 524-525.
- Cox LA. (2011) Risk Analysis of Complex and Uncertain Systems. *Springer Berlin* 129: 943-946.
- Vose. D. (2008) *Risk analysis: a quantitative guide*: Wiley.
- Edwards W, Miles RF and Von Winterfeldt D. (2007) *Advances in decision analysis : from foundations to applications*: Cambridge University Press.
- Kynn M. (2008) The 'heuristics and biases' bias in expert elicitation. *Journal of the Royal Statistical Society* 171: 239-264.
- Pearl J. (2009) *Causality : models, reasoning, and inference*: Cambridge University Press.
- Prelec D. (2004) A Bayesian Truth Serum for Subjective Data. *Science* 306: 462-466.

Index of Authors

Adeyeri, M. K., [133](#), [142](#)
Alencar, M. H., [47](#)
Alkali, B., [1](#)
Alkali, B.M., [179](#)
Almeida, P., [8](#)
Andrews, J.D., [96](#)
Animah, I., [14](#)

Bérenguer, C., [108](#), [168](#)
Bai, F., [114](#)
Bai, Y., [17](#), [77](#)
Bajestani, M. A., [29](#)
Banjevic, D., [29](#), [156](#)
Baocheng, L., [248](#)
Barros, A., [242](#)
Bouacha, Kh., [90](#)
Bull, T., [58](#)

Castro, I. T., [150](#)
Cavalcante, C. A. V., [199](#)
Chatelain, F., [108](#)
Chauvel, C., [35](#)
Chen, H., [41](#)
Chen, Y., [70](#)

Dauxois, J. Y., [35](#)
De Almeida, A. T. , [47](#)
Dhouibi, H., [127](#)
Dinmohammadi,F., [1](#)
Do, P., [168](#)
Doyen, L., [35](#)

Ebenuwa,A. U., [52](#)
Ekpiwhre, E. O., [58](#)

Farrag, M., [179](#)

Gall, A., [168](#)
Gaudoin, O., [35](#)
Gitzel, R., [64](#), [224](#)
Gmati, N., [162](#)
Gong, W., [70](#)
Guo, B., [41](#)
Guo, C., [77](#)
Guo, Y., [235](#)

Hassis, H., [162](#)
Henriques, E., [8](#)
Hu, H., [193](#)
Huynh, K. T., [168](#)

Infante, V., [8](#)

Jia, Y., [77](#)
Jiang, R., [84](#)
Jianmin, Z., [17](#)

Kang, R., [70](#)
Kapodistria, S., [102](#)
Khamel, S., [90](#)
Kilsby, P., [96](#)
Kimura, M., [174](#)
Kuo, W., [121](#)
Kyalosi, Sz., [102](#)

Le, T. T., [108](#)
Lei, P., [114](#)
Li, Z., [114](#)
Lin, M., [248](#)
Liu, B., [121](#)
Liu, F., [193](#)

Maaref, B., [127](#)
Maczey, S., [64](#)
Makinde, O. A., [133](#), [142](#)
Manai, A., [162](#)
Medeiros,C., [47](#)
Mercier, S., [150](#)
Messaoud, H., [127](#)
Montgomery, N., [156](#)
Mordi, O., [58](#)
Mpofu, K., [133](#), [142](#)

Ncib, L., [162](#)
Nguyen, T. P. K. , [168](#)

Ota, S., [174](#)

Qatan, M., [179](#)
Qiwei, H., [17](#)

Ramatsetse, B. I., [133](#), [142](#)
Rauzy, A., [242](#)
Raza, A., [186](#)
Remenyte-Prescott, R., [96](#)
Ren, J. H., [193](#)
Resing, J. A. C., [102](#)
Ribeiro, L. F. A., [199](#)
Rodriguez, T. F. O., [199](#)

Satoglu, S. I., [205](#)
Schmitz, B., [64](#)
Shafiee, M., [14](#)
Simeu-Abazi, Z., [127](#)
Simms, N., [14](#)
Sun, Z. L., [235](#)

Tang, L.C., [217](#)
Tee, K. F., [58](#)
Tee, K. F. , [52](#)
Topcu, O. M., [205](#)
Turrin, S., [64](#), [224](#)

Ulansky, V., [186](#)

Wang, J., [114](#), [235](#)

Wang, W., [211](#)

Wang, X., [217](#)

Wu, S., [64](#), [224](#)

Wu, X., [230](#)

Xie, M., [121](#), [211](#)

Xie, W., [217](#)

Xiong, J., [211](#)

Ye, Z., [217](#)

Yin, M. A., [235](#)

Zhang, W., [70](#)

Zhang, Y., [242](#)

Zhenfei, S., [248](#)

Zhou, C., [179](#)