A CONTROL-LIMIT POLICY AND SOFTWARE FOR CONDITION-BASED MAINTENANCE OPTIMIZATION[1]

D. BANJEVIC, A.K.S. JARDINE, V. MAKIS AND M. ENNIS

Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Canada, M5S 3G8

ABSTRACT

The focus of the paper is the optimization of condition-based maintenance decisions within the contexts of physical asset management. In particular, the analysis of a preventive replacement policy of the control-limit type for a deteriorating system subject to inspections at discrete points of time is presented. Cox's PHM with a Weibull baseline hazard function and time dependent stochastic covariates is used to describe the failure rate of the system. The methods of estimating model parameters and the calculation of the optimal policy are given. The structure of the decision-making software EXAKT is presented. Experience with collecting, preprocessing and using real oil and vibration data is reported.

Keywords: condition-based maintenance, proportional-hazards model, Markov process, cost minimization, decision software

RÉSUMÉ

Dans ce rapport il s'agit de l'optimisation du processus décisionnel par rapport à un programme de l'entretien par surveillance de la condition des machines. Plus spécifiquement on décrit l'analyse d'une politique de maintenance préventive appliqué à un système qui se détériore mais qui est sujet aux inspections aux moments précis. Le modèle par Cox qui si traite aux risques proportionnelles (PHM) comprennant une ligne de base Wiebull ainsi que des co-variants stochastiques est employé dans le bût de décrire le taux de défaults du système. Des méthodes pour effectuer l'estimation des paramètres du modèle ainsi que le calcul de la politque optimale sont présentés. L'architecture du logiciel décisionnel, EXAKT, est décrit. On inclut, également, dans ce rapport, de l'experience sur le collecte, le traitement, et l'usage des données provenant d'un programme d'analyse d'huile et de la vibration.

Mots clés: programme de l'entretien par surveillance de la condition des machines, modelisation des risques proportionnelles (PHM), processus Marcov, minimisation des coûts, logiciel décisionnel.

1. INTRODUCTION

The optimization of decisions in the field of physical asset management (PAM) is an area of increasing interest to management since many PAM decisions are integral to the optimization of supply chain management (SCM) decisions. For example in open-pit mining PAM decisions may consume up to 50% of the annual operating budget. PAM expenditures are also significant in capital intensive industries such as utilities, petrochemical and steel-making. Many PAM decisions are derived through drawing on the methodology of reliability centered maintenance (RCM) with a frequent outcome being the use of condition-based maintenance (CBM). Examples of CBM are the use of information obtained from oil analysis (the spectrometric analysis of metal particles in oil samples regularly taken from an engine's or transmission's lubricating oil), vibration analysis (the spectral analysis of a vibration signal taken at certain positions on rotating machinery), fuel consumption, environmental conditions, etc. In current practice,

[1]Acc. June 2000.

methods for making CBM decisions are based mainly on an engineer's experience and on warning levels for appropriate variables, e.g. for parts per million (ppm) of iron in oil samples. More advanced methods that utilize the condition information for decision making are presented in this paper.

The classical age replacement strategy recommends replacing an item either at failure or when it reaches a certain age, calculated to minimize the expected cost per unit time. The advantage of CBM approach is that it takes into account both the age of the item and its history until the moment of decision making. The condition information can be considered as a vector of covariates, each representing a certain measurement. A convenient method to utilize the covariate vector is to include it in the hazard rate function, using a Proportional-hazards model (PHM), as was proposed by Cox in 1973 (Cox and Oakes (1984)). The PHM and its variants have become one of the most widely used tools in the statistical analysis of the lifetime data in biomedical sciences and reliability. In this paper, a parametric PHM with baseline Weibull hazard function and time dependent stochastic covariates is considered. To calculate the average time to failure, or the average cost associated with a certain replacement policy, it is necessary to introduce a probabilistic model for the covariate process $\{Z(t)\}$. The state space of $\{Z(t)\}$ consists of vectors with coordinates representing specific covariate states. The joint probability distribution model of time to failure T and the process $\{Z(t)\}$ is introduced in Section 2. It is assumed that $(N(t), Z(t))$, where $N(t) = I(T > t)$, is a nonhomogeneous Markov process with a finite state space. The maximum likelihood method is used to estimate the parameters of the PHM and the transition probabilities of the Markov process. Additional details are given in Appendices 1 and 2. Then the estimated statistical model is applied to calculate the decision policy. An obvious extension of the classical age replacement policy is the policy to replace an item either at failure, or when covariates reach some predefined "alarm" states. The level of the "alarm" state may depend on the current age of the item. Combining this method with a cost structure that includes the preventive and failure replacement costs, an optimal policy can be calculated to minimize the expected cost per unit time. If the hazard function is increasing (as it is commonly assumed in research papers) then the optimal policy is very simple, i.e. of the control-limit type (Aven and Bergman (1986), Makis and Jardine (1991)). But in practice, a hazard function that includes covariates as in PHM is often "non-monotonic", at least due to sampling variations, or regular maintenance practices such as oil changes that can affect covariate values. Since the control-limit is applied to the hazard function, this simple policy can be still useful to prevent high risk, even if the hazard function is not strictly increasing. This general case is discussed in Section 3. Relaxation of the "monotonicity" assumption significantly increases the complexity of the calculation. Calculation of the cost function is considered in detail in Appendix 4.

The CBM consortium research group was established in 1995 at the Department of Mechanical and Industrial Engineering, University of Toronto. The goal of the project was to develop software that can assist engineers to optimize decisions in CBM environment. The current development of the software is briefly presented in Section 4. Some experience with collecting, preprocessing and using real oil and vibration data for estimation and modeling is given in Section 5. Proposed research extensions are discussed in Section 6.

2. STATISTICAL MODEL

We consider a replacement model in which an item is replaced with another one "as good as new", either at failure, or at planned replacement. It assumes that item histories are independent and identically distributed random processes. A history includes the information on the item's observed lifetime, censoring information and information on the diagnostic (concomitant) variables collected during the observed lifetime. Let T be the time to failure of the item, and $Z(t) = (Z_1(t), Z_2(t), \ldots, Z_m(t))'$ an m-dimensional covariate process, observed at regular inspections of the item. It is assumed that $Z(t)$ is a right continuous process, with left-hand limits. In practice the coordinates of $Z(t)$ can represent both the external variables (environmental conditions), and internal (diagnostic) variables. The external covariates can affect the time to failure, and the internal variables (such as level of a wear out metal in engine oil) can reflect the current state of the item. We will not consider any formal definition of "external" and "internal" covariates. For an extensive, but rather informal discussion on covariate "types", see Kalbfleisch and Prentice (1980, Ch. 5). We will assume that each covariate is a discrete numerical variable with finite number of values. These values can represent states, such as *new, normal, warning, dangerous*, using figures 0, 1, 2, 3 or midpoints of class intervals for physical measurements, such as for vibration level or ppm (parts per million) of iron in an engine's oil. So, the state space of the process $Z(t)$ is also finite. To introduce a joint distribution of T and $Z(t)$, we will define a right-continuous process $N(t) = I(T > t)$ ($I(\circ)$ being the indicator function) and assume that $(N(t), Z(t))$ is a nonhomogeneous Markov process in a sense that

$$\begin{aligned} &P(T > t, Z(t) = j | T > s,\ Z(s) = i,\ Z(s_{k-1}) = i_{k-1}, \ldots, Z(s_0) = i_0) \\ &\quad = P(T > t, Z(t) = j | T > s,\ Z(s) = i) = P_{ij}(s,t) \qquad (1) \end{aligned}$$

for any $0 \le s_0 < s_1 < \ldots < s_{k-1} < s < t$ and states $i_0, i_1, \ldots, i_{n-1}, i, j$. A necessity for this kind of definition may be in fact that values of $Z(t+\varepsilon)$ for any $\varepsilon > 0$ "... may even not conceptually exist (i.e., covariate processes may be randomly stopped by the corresponding failure times)."(Self and Prentice (1982)). Thus, this definition includes both "external" and "internal" covariates. For simplicity we can assume that $Z(t)$ is available at time points $t = k\Delta$, $\Delta > 0$, $k = 0, 1, 2, \ldots$ (inspection points), and that $Z(t)$ for $k\Delta \le t < (k+1)\Delta$ can be approximated by $Z_k = Z(k\Delta)$. Let

$$p_{ij}(k) = P(Z_{k+1} = j | T > (k+1)\Delta, Z_k = i) \qquad (2)$$

be the (conditional) transition probabilities of the process $\{Z_k\}$. Let also

$$P(T \in (t, t+dt] | T > t, Z(s), s \le t) = h(t, Z(t))dt,$$

i.e. $h(t, Z(t))$ is the hazard function of T. Then we have

$$\begin{aligned} &P(T > k\Delta + x | T > k\Delta,\ Z(s),\ s \le k\Delta) \\ &\quad = \exp\Big\{-\int_{k\Delta}^{k\Delta+x} h(t, Z(t))dt\Big\} = \exp\Big\{-\int_{k\Delta}^{k\Delta+x} h(t, Z_k)dt\Big\},\ 0 \le x \le \Delta, \end{aligned}$$

and also

$$\begin{aligned} p_{ij}(k\Delta, (k+1)\Delta) &= P(T > (k+1)\Delta, Z_{k+1} = j | T > k\Delta, Z_k = i) \\ &= P(T > (k+1)\Delta | T > k\Delta, Z_k = i) \\ &\qquad \times P(Z_{k+1} = j | T > (k+1)\Delta, Z_k = i) \\ &= \exp\Big\{-\int_{k\Delta}^{(k+1)\Delta} h(t, Z(t))dt\Big\} p_{ij}(k). \qquad (3) \end{aligned}$$

If we consider the history $(T, Z) \equiv (T, (Z(s); s \leq T))$, then

$$\begin{aligned} P(T > t, Z(s); s \leq t) &= P(T > t, Z_0, Z_1, Z_2, \ldots, Z_k; k\Delta \leq t < (k+1)\Delta) \\ &= P(T > 0, Z_0)\Big[\prod_{l=0}^{k-1} P_{Z_l Z_{l+1}}(l\Delta, (l+1)\Delta\Big] P_{Z_k Z(t)}(k\Delta, t) \\ &= P(T > 0, Z_0)\exp\Big\{-\int_0^t h(s, Z(s))ds\Big\}\prod_{l=0}^{k-1} P_{Z_l Z_{l+1}}(l). \quad (4) \end{aligned}$$

It should be noted that $S(t; Z) = S(t; Z(s),\ s \leq t) = \exp\Big\{-\int_0^t h(s, Z(s))ds\Big\}$ is frequently confused with the conditional reliability function $R(t; Z) = P(T > t|Z(s), s \leq t)$. If Z is an internal covariate process, $R(t; Z)$ can be meaningless or trivial, taking values 0 and 1 only. If $Z(t)$ is an external process, and T is a "randomized" stopping time (Kebir (1991), Pitman and Speed (1973)), then $R(t; Z)$ exists and $R(t; Z) = S(t; Z)$. Nevertheless, if $h(t, Z(t))$ $p_{ij}(k)$ and $p_i(0) = P(T > 0,\ Z_0 = i)$ do not have common parameters, then $S(t; Z)$ and its derivative $g(t; Z) = h(t, Z(t))S(t; Z)$ can be used to construct the likelihood function and estimate the parameters of the hazard function $h(t, Z(t))$. In either case, it can be considered as a contribution to the partial likelihood (Kalbfleisch and Prentice (1980, Ch. 5), Andersen et al (1993, Ch. 2)). Let $(T_i,\ C_i,\ Z^{(i)}(s);\ s \leq T_i)$, $i = 1, 2, \ldots, n$, be a sample of n independently observed histories, where C_i is the censoring indicator. We assume right-censored data and independent censoring. Then the likelihood of the sample is

$$L(\theta) \propto \prod_{i:C_i=1} h(T_i, Z^{(i)}(T_i); \theta)\prod_j S(T_j, Z^{(j)}; \theta). \quad (5)$$

The parameter θ of the hazard function can be estimated by the maximization of L, or maximization of log-likelihood $l = \log L$. In this paper we consider a parametric PHM with baseline Weibull hazard function as a model for the hazard function. This model is also known as a Weibull parametric regression model. Then

$$\begin{aligned} h(t, Z(t); \beta, \eta, \gamma) &= \frac{\beta}{\eta}\Big(\frac{t}{\eta}\Big)^{\beta-1}\exp\Big\{\sum_1^m \gamma_i Z_i(t)\Big\} \\ &= h_0(t, \beta, \eta)\exp\{\gamma' Z(t)\},\ \beta > 0,\ \mu > 0,\ \gamma' = (\gamma_1, \gamma_2, \ldots, \gamma_m) (6) \end{aligned}$$

Let $0 = t_{i0} < t_{i1} < \ldots < t_{ik_i} = t_i$ be the actual inspection points for the ith history, and $z_j^{(i)} = Z^{(i)}(t_{ij})$ be the observed covariate values, $j = 0, 1, \ldots, k_i$, $i = 1, 2, \ldots, n$. The inspection times are usually not equally spaced, but it can be ignored if the intervals between successive inspections are not very variable. The factor $S(t, z^{(i)})$ can be calculated as follows

$$\begin{aligned} S(t, z^{(i)}) &= \exp\Big\{-\int_0^{t_i} h_0(x)\exp\{\gamma' z^{(i)}(x)\}dx\Big\} \\ &= \exp\Big\{-\sum_{j=0}^{k_i-1}\int_{t_{ij}}^{t_{i(j+1)}} \exp\{\gamma' z^{(i)}(x)\}d(x/\eta)^\beta\Big\} \\ &= \exp\Big\{-\sum_{j=0}^{k_i-1}\exp\{\gamma' z_j^{(i)}\}\Big[\big(t_{i(j+1)}/\eta\big)^\beta - (t_{ij}/\eta)^\beta\Big]\Big\}, \quad (7) \end{aligned}$$

using that $z^{(i)}(t) = z_j^{(i)}$ for $t_{ij} \le t < t_{i(j+1)}$. Some other approximations can also be used, such as $z^{(j)}(t) = (z_j^{(i)} + z_{j+1}^{(i)})/2$ for $t_{ij} \le t < t_{i(j+1)}$. The latter one could be more suitable for "non-monotonic" covariates. If there is no covariate measurement $z_{k_i}^{(i)} = z^{(i)}(t_i)$ at the event time, some extrapolation is necessary. For "monotonic" covariates some polynomial trending, such as linear, can be used, but for "non-monotonic" covariates this can often produce an unrealistic result (very high, or very low value). In general, a simple method to use the previous measurement $z^{(i)}(t_i) = z^{(i)}(t_{i(k_i-1)}) = z_{k_i-1}^{(i)}$ can be recommended. In either case

$$h(t, z^{(i)}(t_i)) = \frac{\beta}{\eta}\left(\frac{t_i}{\eta}\right)^{\beta-1} \exp\{\gamma' z^{(i)}(t_i)\}, \tag{8}$$

and then the total log-likelihood is

$$l = l(\beta,\eta,\gamma) \propto r\ln(\beta/\eta) + (\beta-1)\sum_i \ln(t_i/\eta) + \sum_i \gamma' z^{(i)}(t_i)$$
$$-\sum_i \int_0^{t_i} \exp\{\gamma' z^{(i)}(x)\} d(x/\eta)^\beta, \tag{9}$$

where r is the number of failures. A maximization technique can be used to estimate the parameters β, η, γ, e.g. BFGS Quasi-Newton method (Press at al (1992)).

The method of maximum likelihood can be also used to estimate the transition probabilities of the Markov process model. If the inspection instants are (at least roughly) equally spaced, then the estimator for the transition probability in (2) is

$$\hat{p}_{ij}(k) = \frac{n_{ij}(k)}{n_{i\bullet}(k)}, \quad k = 1, 2, \ldots, \tag{10}$$

where $n_{i\bullet}(k) = \sum_j n_{ij}(k)$, and $n_{ij}(k)$ is the number of one-step transitions $i \to j$ at the time $k\Delta$ in the sample (Basawa and Rao (1980)). In practice, a large sample size (many inspections) would be necessary to obtain good estimates of $p_{ij}(k)$ for each k. It is more convenient to make a partition of the time range into L intervals, $0 = l_0 < l_1 < \ldots < l_L = \infty$, and assume that the process is homogeneous within each of these intervals, i.e. $p_{ij}(k) = p_{ij}(l_{r-1})$, $l_{r-1} \le k < l_r$, $r = 1, 2, \ldots, L$, and then to estimate $p_{ij}(l_{r-1})$ by

$$\hat{p}_{ij}(l_{r-1}) = \frac{\tilde{n}_{ij}(l_{r-1})}{\tilde{n}_{i\bullet}(l_{r-1})}, \tag{11}$$

where $\tilde{n}_{ij}(l_{r-1}) = \sum_{l_{r-1} \le k < l_r} n_{ij}(k)$ and $\tilde{n}_{i\bullet}(l_{r-1}) = \sum_j \tilde{n}_{ij}(l_{r-1})$.

With variable, but still not long intervals between successive inspections, it is more convenient to estimate first the transition rates and then calculate the transition probabilities. Consider for now the process $\{Z(t)\}$ without approximation $\{Z_k\}$. Assume that $Z(t)$ has constant *conditional* transition intensities within given partition of time, $0 = s_0 < s_1 < \ldots < s_L = \infty$, i.e. if $p_{ij}(t|s) = P(Z(t) = j|T > t,\ Z(s) = i)$, then $p_{ij}(s + ds|s) = \delta_{ij} + \lambda_{ij}^{(l)} ds$, $s_l \le s \le t < s_{l+1}$, where δ_{ij} is Kronecker δ symbol, and $\sum_j \lambda_{ij}^{(l)} = 0$. Then the transition rates $\lambda_{ij}^{(l)}$ can be estimated using occurrence/exposure rates

$$\hat{\lambda}_{ij}^{(l)} = \frac{n_{ij}^{(l)}}{A_i^{(l)}},\ i \ne j, \quad \hat{\lambda}_{ii}^{(l)} = -\sum_{j \ne i} \lambda_{ij}^{(l)}, \tag{12}$$

where $n_{ij}^{(l)}$ is the number of all transitions $i \to j$ occurred over the interval $[s_l, s_{l+1})$ in the sample, and $A_i^{(l)}$ is the total length of time that the state i is occupied over the interval $[s_l, s_{l+1})$ in the sample. Note that this method can be applied in practice if inspections are frequent enough to cover almost all transitions in the observed period, as it would be if the transitions occur just before the inspection instants. For the proof, see Appendix 1. Note, however this is not an estimation for an ordinary Markov process, as given in Basawa and Rao (1980, Ch. 8), because here we deal with the transition probabilities conditioned on T. For an alternative nonparametric approach to the estimation of the transition probabilities using the theory of counting processes, see Keiding and Andersen (1989), or Andersen et al (1993). If the covariate process $\{Z(t)\}$ is an external Markov process, then the transition probability matrix can be calculated by

$$P^{(l)}(x) = \exp(\Lambda^{(l)}x) = \sum_{i=0}^{\infty} \frac{(\Lambda^{(l)}x)^i}{i!}, \quad 0 \le x < s_{l+1} - s_l, \tag{13}$$

where $P^{(l)}(x) = (p_{ij}^{(l)}(s+x|s))$, $s_l \le s \le s+x < s_{l+1}$, and $\Lambda^{(l)} = (\lambda_{ij}^{(l)})$. In general, (13) can be used as an approximation to $P^{(l)}(x)$ for small x, in practice not greater than the average interval between successive inspections. For further details, see Appendix 2.

3. OPTIMAL REPLACEMENT POLICY

The model for CBM described in Section 2 assumes inspections of an item at fixed intervals of time, and a decision policy as a rule for replacement or leaving the item in operation until the next decision opportunity. The rule depends on the age of the item and the inspection results. The item is always replaced at failure. An optimal rule is selected to minimize the average replacement costs per unit time due to preventive and failure replacements over a long time horizon. Let $C_p = C$ be the preventive replacement cost, and $C_f = C + K$ be the failure replacement cost, per one replacement. These costs are assumed fixed and equal for all replacements. Assume also that the preventive replacement can be planned at any moment. Let $d > 0$ be a "control-limit" value, and $T_d = T_d(Z(s); s \ge 0)$ a stopping rule of the form

$$T_d = \inf\{t \ge 0 \colon Kh(t, Z(t)) \ge d\}, \tag{14}$$

i.e. if $T_d < T$, perform the preventive replacement at T_d, and if $T_d \ge T$, perform the failure replacement at T. Let also

$$Q(d) = P(T_d \ge T), \; W(d) = E(\min\{T_d, T\}). \tag{15}$$

The expected cost per unit time is then

$$\Phi(d) = \frac{C(1 - Q(d)) + (C+K)Q(d)}{W(d)} = \frac{C + KQ(d)}{W(d)}. \tag{16}$$

The optimal control-limit replacement rule $T^* = T_{d^*}$ is defined by d^* for which $\Phi(d^*) = \min_{d>0} \Phi(d)$. If the hazard function $h(t, Z(t))$ is non-decreasing with time, the rule T^* is proved to be optimal in general, i.e. it is the best possible replacement policy (Aven and Bergman (1986), Makis and Jardine (1991)). Also, $\Phi(d^*) = d^*$ and the optimal level d^* can be found using the fixed-point iteration procedure $d_n = \Phi(d_{n-1})$, $n = 1, 2, \ldots, d_0 > 0$ arbitrary. A good choice is $d_0 = \Phi(\infty) = (C+K)/E(T)$, to start with the policy to replace only at failure. The iteration procedure usually requires only a few steps, but the calculation of the function $\Phi(d)$ is complicated and can take a relatively long time if the number of states for the process $\{Z(t)\}$ is large. This can be the case even with few covariates. We will consider this calculation later.

The assumption about increasing hazard is theoretically very appealing, but rarely met in practice. In our statistical model covariates are considered as random, and their stochastic behavior is directed by the transition probabilities (2). In practice, some regular maintenance procedure (not considered as a major overhaul), such as oil change, or regular replacement of some minor component, can slightly improve the equipment state and reduce the hazard rate, even if it has an increasing trend. If the covariates can decrease, at least due to sampling variations, this can also make the hazard function "non-monotonic". In that case the optimal control-limit rule T^* is not optimal in general. But if the hazard rate has an increasing trend, such as in our model (6) for $\beta > 1$, T^* can be "near to optimal". Also, it is hard to find some practical decision rule that would be optimal in the general case, with non-monotonic hazard function. We will not assume that hazard function is increasing, but we will again consider the control-limit rule, with a slight modification, and try to find the optimal level d^*. Let in (6) $\beta < 1$. Then $h(0, Z(0)) = \infty$, and $T_d \equiv 0$ for all $d > 0$, i.e. the rule would require immediate replacement at 0. So, the control-limit policy cannot be applied in general without some restrictions. Assume that $h(t, Z(t)) < \infty$ for all $t > 0$. We will introduce the minimal preventive replacement time $t_R \geq 0$ and define the stopping rule

$$T_{t_R,d} = \inf\{t \geq t_R\colon Kh(t, Z(t)) \geq d\} = \max\{t_R, T_d\}. \tag{17}$$

If $Kh(t, Z(t)) < d$ for all $t \geq t_R$, then $T_{t_R,d} = \infty$. So, $T_{t_R,d} \geq t_R$, i.e. no preventive replacement is planned before t_R, regardless of the covariate process $Z(t)$. The time before t_R can be considered as a "run-in" period, or a "warranty" period. Let K_R be the failure replacement cost for a failure in the period $[0, t_R]$ (usually $C \leq K_R \leq C+K$). Let $Q_{t_R}(d) = P(T_{t_R,d} \geq T)$, $W_{t_R}(d) = E(\min\{T_{t_R,d}, T\})$. Then the expected cost per unit time is

$$\Phi_{t_R}(d) = \frac{C + KQ_{t_R}(d) - (C + K - K_R)P(T \leq t_R)}{W_{t_R}(d)}, \tag{18}$$

(see Appendix 3). Calculation of the cost function (18) is more complex than for the "increasing" case, and the optimal level d^* cannot be found by just using the iteration procedure, but can be obtained using direct minimization. The cost function can be calculated at a set of points d_i, and then d^* can be approximated by d_0^* for which $\Phi_{t_R}(d_0^*) = \min_i \Phi_{t_R}(d_i)$, or using an interpolation. An additional problem is that $\Phi_{t_R}(d)$ is not in general a continuous function, as it is for the "monotonic" case. For example, if in the Weibull regression model $\beta \leq 1$, then $\Phi_{t_R}(d)$ is a step function. The calculation for lot of points d_i in such cases is sometimes necessary. The calculation of the function $\Phi_{t_R}(d)$ is considered in detail in Appendix 4. It is shown that the calculation is convenient for a stopping rule of "multiplicative" form (see Notes in Appendix 4).

Once the optimal threshold level d^* is calculated, then the optimal replacement rule is to replace the item at the first moment $t \geq t_R$ such that $h(t, Z(t)) \geq d^*/K$, or in the Weibull regression model case, when $(\beta/\eta)(t/\eta)^{\beta-1} \exp\left(\sum \gamma_i z_i(t)\right) \geq d^*/K$. A more intuitive form of this decision rule is

$$T^* = \min\left\{t \geq t_R\colon Z^C(t) \equiv \sum \gamma_i Z_i(t) \geq \delta^* - (\beta - 1)\ln t\right\} \tag{19}$$

$\delta^* = \ln(\eta^\beta d^*/(\beta K))$. The function $g(t) = \delta^* - (\beta - 1)\ln t$ can be considered as a "warning level" function, applied to an "overall" covariate value $Z^C(t)$. If $\beta > 1$, $g(t)$ is a strictly decreasing function (see Figure 1.(a)). If $\beta = 1$, $g(t) = \delta^*$ is a constant "warning" level, as is usually applied in practice (Figure 1.(b)). If $\beta < 1$, then $g(t) = \delta^* + (1-\beta)\ln t$ is an increasing function. In that case it is obvious that high values of $Z^C(t)$ are more important sooner than later in the equipment life (Figure 1. (c)). An interesting feature of the "warning-level" function $g(t)$ is that its shape is not affected by the change of the replacement costs C_f and C_p, it is just translated up or down, because only the constant δ^* is affected by new d^*. It appeared that δ^* is not very sensitive to mild changes in the cost ratio C_f/C_p, but it is hard to give an accurate estimation. This has an important application in practice: only rough estimation of the cost ratio is necessary to get the decision policy close to optimal.

We will also incorporate regular preventive maintenance practice in the model. Assume that the regular maintenance affects the covariates at given intervals of time by forcing covariate values to initial or some predetermined states. Let the time between successive regular maintenance actions be t_M, $t_M \leq \infty$, and i_M is the state "at maintenance" (just after the maintenance) to which the covariates will be returned after the action. The state i_M can also depend on the age of the item when action is taken. For simplicity, let i_M be time-independent. Let $\{\tilde{Z}(t)\}$ be the process $\{Z(t)\}$ "affected" by the maintenance. Then $\tilde{Z}(t)$ is a Markov process with $\tilde{Z}(kt_M) = i_M$, with probability one, and the following transition probabilities between maintenance points:

$$\begin{aligned}\tilde{P}_{ij}(s,t) &= P(T > i, \tilde{Z}(t) = j | T > s, \tilde{Z}(s) = i) \\ &= P_{ij}(s,t), \quad kt_M \leq s < t < (k+1)t_M, \\ \tilde{P}_{ii_M}(s,(k+1)t_M) &= P(T > (k+1)t_M | T > s, \tilde{Z}(s) = i) = P_{i\bullet}(s,(k+1)t_M), \\ \tilde{P}_{ij}(s,(k+1)t_M) &= 0, \quad j \neq i_M,\ kt_M \leq s < (k+1)t_M, \end{aligned} \tag{20}$$

for $k = 0, 1, 2, \ldots,$, where $P_{ij}(s,t)$ is defined in (1), and $P_{i\bullet}(s,t) = \sum_j P_{ij}(s,t)$. Extension beyond the maintenance points is obvious. In an example with engine oil data, regular oil samples were taken and analyzed every 200 hours, and oil and filter changes were performed every 600 hours. Data showed an increasing trend (with variations) between oil changes, so that the oil change information was important for a proper analysis and interpretation of measurements. More discussion on related problems is given in Section 5. If the data incorporates regular maintenance information, this should be used to properly estimate the transition probabilities of $\tilde{Z}(t)$. Only the transitions between two successive maintenance points kt_M and $(k+1)t_M$, $k = 0, 1, 2, \ldots$ should be considered in (10-12) for the estimation of (20). Some enhancements of this "preventive repair" model are possible, such as the "state at maintenance" i_M is random, depending on the history of the process $\{\tilde{Z}(t)\}$ until the repair, but this could make the model more complicate to estimate. For other possible enhancements of the model, see Section 6. A critical discussion on the complexity and application of theoretical models in maintenance practice has been recently given by Scarf (1997).

4. STRUCTURE AND DEVELOPMENT OF THE SOFTWARE FOR CBM

The CBM Laboratory was established in 1995 at the Department of Mechanical and Industrial Engineering, University of Toronto, to develop decision-making software for CBM. The design of the software is based on the replacement policy model introduced in Sections 2 and 3. After five years of development, the Versions 1.0 and 2.0 of the software called EXAKT, were released to Consortium members. We will briefly present the structure of EXAKT and its main options. A more detailed overview of Version 1.0

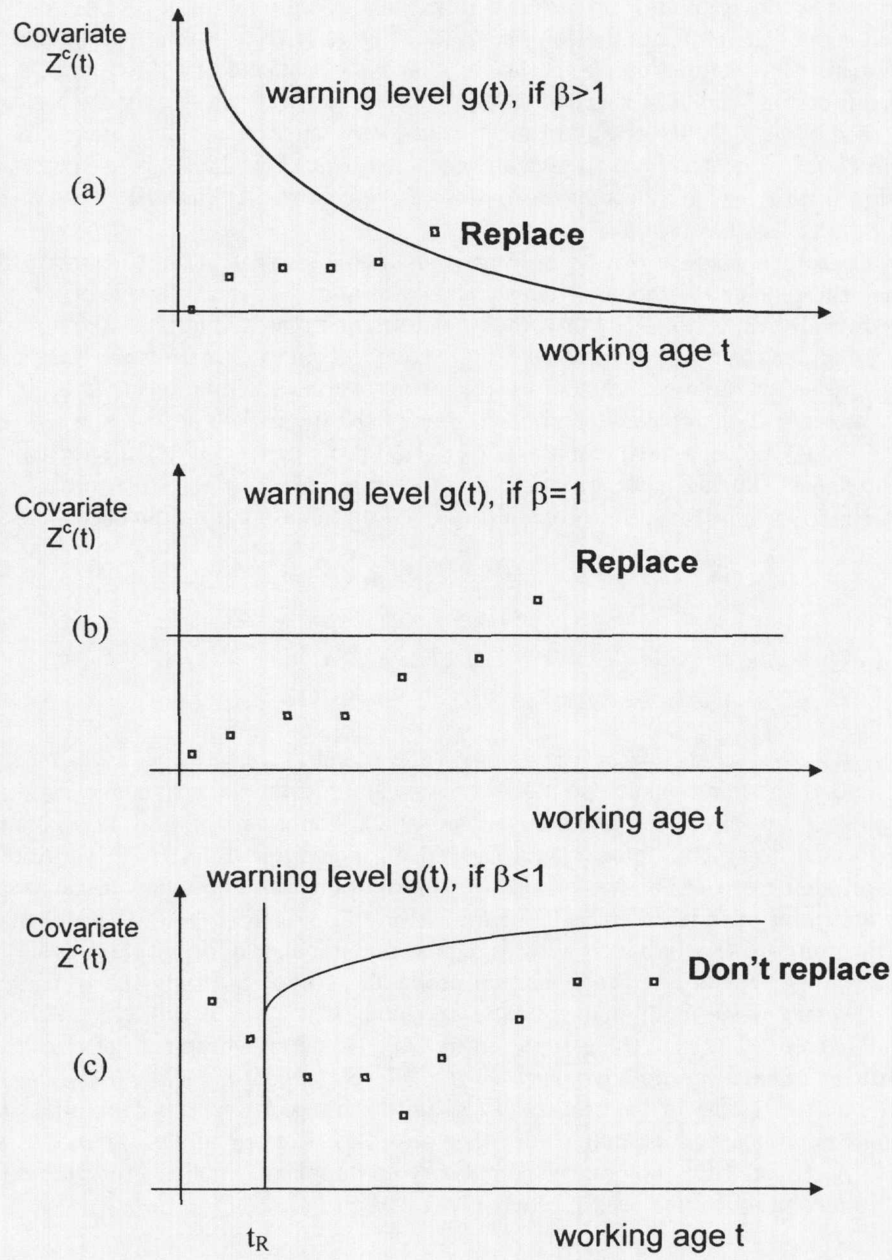


Figure 1: Optimal replacement policy

is given by Jardine et al (1998). The software is being developed in such a way that the user can:

1. Create a convenient database by extracting the event and condition (inspection) information from external databases.

2. Perform data analysis and preprocessing, using graphical and statistical analysis.

3. Estimate parameters of the PHM and Markov process model. The model can be evaluated using Cox-generalized residuals and a variety of tests (Wald test, Log-likelihood test, Kolmogorov-Smirnov test, X^2 test for independence of covariates and for homogeneity of the Markov process).

4. Compute and save the optimal replacement policy. To cover the situation in which condition information is not available for decision making, simple Age and Block replacement policies are also included.

5. Make decisions for current records whenever it is required, using the already saved decision model. The decision recommendation (*replace/don't replace*, and some other useful information, such as the expected "remaining life") is stored in the user's database for further utilization.

EXAKT also provides procedures for the checking, correction and transformation of data. A simplified programming language has been developed and included in EXAKT to help the user analyze the data using graphical methods and a number of statistical operations. The user can also generate new variables by transforming the original data, applying mathematical functions such as *log* or *exp*, linear combinations of variables, or by calculating rates, cumulatives or different kind of trends. Time component can be also included in these transformations: the user can apply different transformations to the same variable in different time intervals, e.g. to fit the PHM with time-dependent coefficients (such as to use piecewise constant functions $\beta_i(t)$, see Murphy and Sen (1991), Marzec and Marzec (1997)).

5. EXPERIENCE WITH REAL DATA

5.1 Data Collection

In this section we will briefly present our experience with real data. Numerous shortcomings in data collection practices were experienced while preparing and analyzing data from collaborating companies, such as:

1. Improperly organized databases.

2. In many cases, only the calendar age is recorded, and not the real usage time (working age) of the component. If the working age is recorded, there are often inconsistencies with the calendar age that have to be cleared up.

3. Usually there are no records on maintenance done during a component's lifetime.

4. If maintenance records are available (often in a form of hand-written workorders, which is close to being useless), there is no proper repair type classification in a form of codes. This could be very convenient for possible use of a model for repairable systems.

5. The cause of replacement or maintenance is usually not recorded, i.e. whether an action was taken preventively, or because of failure.

6. Inspection data is sometimes missing because either it was deleted after a certain period of time, or it was not stored regularly in the database. This problem becomes now less significant with increasing utilization of more powerful computerized maintenance management systems.

7. Inspections are not done regularly, or they are repeated more often when some problem is suspected.

An example of these kind of problems encountered with data requirements and their collection related to the operational decision support system PROMPT can be found in Dekker and Scarf, (1998, Section 3). Similar problems in applying optimization models are also discussed by Dekker (1996, Section 5).

5.2 Modeling

Here we will give some tips in the modeling with oil analysis data for mobile equipment, but they are not uncommon with other types of analyses, such as vibration monitoring.

1. When metal levels are plotted against working age, there is often a "run-in" period for new equipment (such as an engine) when the metal levels are extraordinarily high. These values are not very indicative for the equipment health, and can be replaced by some average values more informative of the metals' long-run behavior. An advanced method suggested by Tibshirani (1988), can be used to standardize these values and extract possible valuable information. The method appeared successful, but not very simple to implement as a regular procedure.

2. Oil analysis data exhibits quite a large amount of variation. As metal particles indicating wear accumulate in oil, levels of these metals should increase with oil age. In practice, a decrease in the metal levels can be often seen, especially in metals that do not show much wear, even with the same oil (for mobile equipment it is usual to have only one or two readings before the oil is changed). This was one of the reasons why we had to consider the "non-monotonic" covariates.

3. It is recommended by technicians to use rates (metal level ÷ oil age) to indicate metal wear. An increase in the rate of wear should indicate failure warning. Unfortunately, miss-specified oil ages - especially when very small - combined with large variations, have a very detrimental effect on rates since they give rise to many spuriously high rates. This problem is also mentioned in Jardine et al (1989). Toms (1998) uses a simple adaptive trend algorithm as an approach to the solution of this problem.

4. As an alternative, one can also consider cumulative metal levels. The function of cumulative levels for a metal would be proportional to the total volume of that metal lost to wear and may therefore be related to the lifetime of the unit, as considered in a different context by Ansell and Phillips (1989). However, the cumulative metal level is unlikely to be sensitive to a sudden change in rate.

5. If replacements of components and minor repairs are not recorded, the metal levels will appear to go up and down for no apparent reason. This may account for cases we have seen where there are no increasing patterns in the average metal levels over the long term. This will decrease the significance of related "regression" coefficient γ_i in PHM. The problem of minor repairs is also related to the decision on how to build the model and possible inclusion of the regular maintenance interval, as discussed in Section 3.

5.3 Case Study

We will give a brief overview of a case study that refers to a shear-pump bearings in a food processing plant. Diagnostic data was obtained from vibration measurements on the bearing, in axial, horizontal and vertical directions. Time domain data is automatically transferred to frequency domain. For each direction seven measurements are obtained: velocity spectrum in five frequency bands, overall velocity, and overall acceleration, which totals in 21 covariates. 25 histories were available, including 13 ended by failures and 12 ended by preventive suspensions. The average history duration, in this case also the average replacement time, was 221 days. 165 inspections were performed during the study period, with the average time between inspections being roughly 35 days. After statistical analysis and PHM estimation procedure, three covariates were finally included in the model: two velocity bands in the axial direction, and one velocity band in the vertical direction. Time (age of bearing) appeared as significant: the estimate of the shape parameter $\hat{\beta} = 4.992$, with the standard error $\hat{\sigma}(\hat{\beta}) = 1.173$. Test model fit was acceptable. These results are consistent with technicians experience, except they expected one or two more covariates to appear in the model. For each covariate in the model, certain covariate states were established using technical standards, with practical interpretation as: "very smooth", "smooth", "rough", "very rough", and the Markov process model was estimated. We will not go into further details of the estimation procedure.

The optimal control-limit decision policy was calculated for the cost ratio $C_p\colon C_f = 1\colon 9$, as suggested by the company. No regular maintenance interval t_M, or minimal preventive replacement time t_R were included. The calculated cost of the optimal policy was \$16.04/day, and of the policy to replace only at failure was \$74.79/day. The expected time between replacements for the optimal policy was 173 days, and for the policy to replace only at failure was 283.32 days. The former value can be compared with the company's actual mean replacement time of 221 days, and the latter with a non-parametric estimate of the mean time to failure of 297 days (using Kaplan-Meier product-limit estimator for right censored data). The company's actual cost was \$55.02/day (calculated from the number of failures and suspensions), which is better than to replace only at failure, but much worse than the optimal policy.

To check whether the calculated optimal policy is reasonable, different methods can be applied afterwards to the analyzed histories. The method applied here is conservative, and is not straightforward, particularly with a small sample size, and will be discussed in a future paper. As a result, 7 failures, 17 suspensions and one undecided case would be obtained, if the optimal policy were followed, instead of the actual 13 failures and 12 suspensions. The "realized" optimal cost would then be \$37.37/day with an average replacement time of 209.71 days. The "realized" optimal cost is larger than the theoretical optimal cost, but still smaller than the actual cost. We think the company's real average cost/day is higher than the calculated \$55.02/day, because some of the histories that were reported as suspended preventively, are actually temporary suspended (calendar suspensions) and it is still unknown whether they will end with failure, or suspension. Also, the difference between the theoretical optimal cost, and the "realized" optimal cost could be explained due to the small sample size, and even more importantly because of very irregular inspections sometimes used by the company. These irregular inspections increased the time between replacements, but also increased the number of failures. It should be noted that the company considers a very high vibration level also as a failure, and this was included in the analyzed data set. In 5 out of 7 "unpredicted" failures, the vibration level was well above the "warning limit" function (see (19) and Figure 1 (a)) at the last inspection. Only in two "unpredicted" cases was the vibration level at failure below the "warning limit".

One can question testing the decision model with the same data used for its calculation, but the test is useful at least as a "crude" check. Obviously, with more or new histories available, there are other possibilities, for example, one group of randomly selected histories could be used to build the model, and another one to test the model.

6. RESEARCH DIRECTIONS AND ENHANCEMENTS

Possible research extensions of the decision model presented in Sections 2 and 3, are in the area of repairable systems, and optimization of inspection and maintenance intervals. A combination of repairable systems with multi-component systems and the PHM is of great importance because of fairly complicated maintenance practices for complex systems and fast growing interest in CBM. It can include different failure modes, partial repairs or opportunistic maintenance and other preventive maintenance practices. The current maintenance practice often produces histories with almost no failures. An attempt to overcome this problem is made by introducing two failure modes, Makis (1995), one related to catastrophic failures, and the other related to reaching some "warning" level of a diagnostic variable. Of theoretical and practical interest is also how to deal with partial or incomplete data, e.g. with histories that can report failures or suspensions with segments of the inspection records missing. Incomplete data could be a regular situation: some variables intentionally are not recorded at every inspection. This cannot be considered as an ordinary "missing data" problem. The question is not only how to fit the statistical model, but also how to include this partial information in the decision making. Inclusion of some more advanced non-constant cost models can be also considered, particularly if a multi-failure, or multi-component model is more realistic, but only if an appropriate cost data is possible to collect. As mentioned by Scarf, regarding extensions of the maintenance models (1997): "Often, sufficient data are not available to consider complex models; even if data are available, the maintenance policies implied by complex models [...] may be difficult to implement in practice."

7. CONCLUDING REMARKS

The growing application of condition monitoring techniques in maintenance decision making has produced a challenge for researchers to develop appropriate more advanced decision models. Proportional-hazards modeling of the risk estimation of equipment failure combined with economic factors is found to be a useful method of utilizing the condition-monitoring information. Close collaboration with supporting companies has become a rich source of industry driven research topics, as well as a base for development and testing of the software for condition-based maintenance.

ACKNOWLEDGMENTS

Appreciation is expressed to the following participants in the Condition-Based Maintenance Consortium: Materials and Manufacturing Ontario, The Natural Sciences and Engineering Research Council of Canada (grant # 661-215544/98 CRDPJ), ALCOA, Campbell Soup Company Ltd., Department of National Defence, Dofasco Inc., Hong Kong Mass Transit Railway Corporation, Oliver Interactive Inc., PricewaterhouseCoopers, and Syncrude Canada Inc. We also appreciate Thampu Joseph's role in developing the case study.

APPENDIX 1.

Estimation of the transition probability matrix

Let $P_{ij}(s,t) = P(T > t, Z(t) = j|T > s,\ Z(s) = i)$ and $P(s,t) = [P_{ij}(s,t)]$. Note that

$P(s,t)$ is not a stochastic matrix, because $\sum_j P_{ij}(s,t) = P(T > t|T > s,\ Z(s) = i) \le 1$. By (1) $P(s,t) = P(s,u)P(u,t)$, $s \le u \le t$, where is assumed that $P(s,s) = \lim_{t \downarrow s} P(s,t) = I$. Let $\frac{\partial}{\partial t}P(s,t)|_{t=s} = \alpha(s)$ exists. Then $\frac{\partial}{\partial t}P(s,t) = P(s,t)\alpha(t)$. Let $\lambda_{ij}(s) = \frac{\partial}{\partial t}p_{ij}(t|s)|_{t=s} = \frac{\partial}{\partial t}P(Z(t) = j|T > t,\ Z(s) = i)|_{t=s} = \lambda_{ij}^{(l)}$, $\lambda_i^{(l)} = -\lambda_{ii}^{(l)} = \sum_{j \ne i}\lambda_{ij}^{(l)}$, for $s_l \le s < s_{l+1}$, and $\Lambda^{(l)} = (\lambda_{ij}^{(l)})$. Let $D(s) = \mathrm{diag}[h(s,i)]_i$. It can be shown that $\alpha(s) = \Lambda^{(l)} - D(s)$, i.e. $\alpha_{ij}(s) = \lambda_{ij}^{(l)} - \delta_{ij}h(s,i)$, $s_l \le s < s_{l+1}$. Let $\alpha_i(s) = -\alpha_{ii}(s) = \lambda_i^{(l)} + h(s,i)$, $s_l \le s < s_{l+1}$, and $\alpha_{iF}(s) = P(T \le s+ds|T > s,\ Z(s) = i)/ds = h(s,i)$. Let also $p_i(0) = P(T > 0,\ Z(0) = i)$.

Let $(t,\delta,z) = (t,\delta,(z(s),\ 0 \le s \le t))$ be an observed sample history, where $T = t$ if $\delta = 1$ and $T > t$ if $\delta = 0$, and the complete observation of the covariate process $Z(s)$ is $z(s)$, for $0 \le s \le t$. Let $0 = x_0 < x_1 < x_2 < \ldots < x_k < t$ be the time points at which $z(t)$ changed the state, and $z(x_j) = i_j$, $j = 0,1,\ldots,k$, $z(t) = z(x_k) = i_k \cdot z(s)$ is assumed right continuous, with left limits. Let us note that with probability one there is no jump of $Z(t)$ at T. Then the likelihood of (t,δ,z) has the form proportional to

$$
\begin{aligned}
&p_{i_0}(0) \times \prod_{j=0}^{k-1}\Big[\prod_{x_j \le s < x_{j+1}}(1-\alpha_{i_j}(s)ds)\alpha_{i_j i_{j+1}}(x_{j+1})\Big] \times \prod_{x_k \le s \le t}(1-\alpha_{i_k}(s)ds) \\
&\quad \times [\alpha_{i_k F}(t)]^\delta \\
&= p_{i_0}(0) \times \prod_{j=0}^{k-1}\exp\Big\{-\int_{x_j}^{x_{j+1}}\alpha_{i_j}(s)ds\Big\} \times \prod_{j=0}^{k-1}\alpha_{i_j i_{j+1}}(x_{j+1}) \times \exp\Big\{-\int_{x_k}^{t}\alpha_{i_k}(s)ds\Big\} \\
&\quad \times [h(t,i_k)]^\delta \\
&= p_{i_0}(0) \times \exp\Big\{-\int_0^t \alpha_{z(s)}(s)ds\Big\} \times \prod_{l,i,j}(\lambda_{ij}^{(l)})^{n_{ij}^{(l)}} \times [h(t,i_k)]^\delta \\
&= p_{i_0}(0) \times \exp\Big\{-\int_0^t [h(s,z(s)) + \lambda_{z(s)}(s)]ds\Big\} \times \prod_{l,i,j}(\lambda_{ij}^{(l)})^{n_{ij}^{(l)}} \times [h(t,i_k)]^\delta \\
&= p_{i_0}(0) \times \exp\Big\{-\int_0^t h(s,z(s))ds\Big\} \times [h(t,i_k)]^\delta \times \exp\Big\{-\sum_{l,i}\lambda_i^{(l)}a_i^{(l)}\Big\} \\
&\quad \times \prod_{l,i,j}(\lambda_{ij}^{(l)})^{n_{ij}^{(l)}},
\end{aligned}
$$

where $n_{ij}^{(l)}$ is the number of all transitions $i \to j$ occurred over the interval $[s_l, s_{l+1})$ in the history, and $a_i^{(l)}$ is the total length of time that state i is occupied over the interval $[s_l, s_{l+1})$ in the history. The total likelihood for a sample of independent histories can be obtained as the product of the terms of the previous form. If the initial probability $p_i(0)$ and the hazard function $h(s,i)$ do not depend on $\lambda_{ij}^{(l)}$, then the ML estimates (12) of the form occurrence/exposure rates can be easily obtained. The method of counting processes can be also used to obtain the likelihood of the sample (see Andersen et al (1993, III.1.2)).

APPENDIX 2.

Calculation of the transition probability matrix

The transition probability matrix $P(s,t)$ (see Appendix 1) can be expressed as a product integral

$$P(s,t) = \prod_{(s,t]}(I + (\Lambda^{(l)} - D(x))dx), \quad s_l \le s < t \le s_{l+1}, \quad l = 1,2,\ldots \tag{21}$$

The product integral (21) can be computed as a finite product

$$\prod_j(I + (\Lambda^{(l)} - D(x_j))\Delta x_j), \quad \Delta x_j = x_{j+1} - x_j, \quad s = x_0 < x_1, < \ldots < x_n = t, \tag{22}$$

for $\varepsilon = \max\{\Delta x_j\}$ sufficiently small, at least to provide that diagonal elements of $(I + (\Lambda^{(l)} - D(x_j))\Delta x_j)$ are nonnegative. For an overview of product integration see Andersen et al (1993), or Gill and Johansen (1990). An alternative method to the computation of $P(s,t)$ is the following. Let for a quadratic matrix $B = (b_{ij})$, $\exp(Bx) = \sum_0^\infty (Bx)^k/k! = I + Bx + R_B(x)$, where $\|R_B(x)\| \leq x^2(\|B\|^2/2)\exp\{\|B\|x\}$, $x \geq 0$, and $\|B\|$ is a matrix norm, e.g. $\|B\| = \max_i \sum_j |b_{ij}|$. The matrix $I + (\Lambda^{(l)} - D(x_j))\Delta x_j$ can be approximated by $\exp(-D(x_j)\Delta x_j)\exp(\Lambda^{(l)}\Delta x_j)$. Then, using Duhamel's equation (Andersen et al (1993)),

$$\left\| \prod_j (I + (\Lambda^{(l)} - D(x_j))\Delta x_j) - \prod_j \exp(-D(x_j)\Delta x_j)\exp(\Lambda^{(l)}\Delta x_j) \right\|$$
$$\leq \sum_j \left\| (I + (\Lambda^{(l)} - D(x_j))\Delta x_j) - \exp(-D(x_j)\Delta x_j)\exp(\Lambda^{(l)}\Delta x_j) \right\| \leq \varepsilon^2 C(\varepsilon),$$

where $C(\varepsilon)$ is bounded for $\varepsilon \leq \varepsilon_0$, and ε_0 is selected such that $\max_j \|I + (\Lambda^{(l)} - D(x_j))\Delta x_j\| \leq 1$. If we use an approximation $\exp(-D(x_j)\Delta x_j) \approx \exp\Big(- \int_{x_j}^{x_{j+1}} D(x)dx\Big) = \mathrm{diag}\Big[\exp\Big\{ - \int_{x_j}^{x_{j+1}} h(x,i)dx\Big\}\Big]$, then also

$$P(s,t) \approx \prod_j \exp\Big(- \int_{x_j}^{x_{j+1}} D(x)dx\Big) \exp(\Delta^{(l)}\Delta x_j), \quad s_l \leq s < t \leq s_{\ell+1},$$

as is basically introduced in (3).

APPENDIX 3
Cost function

$\Phi_{t_R}(d)$ =expected cost of one replacement cycle/expected time of one replacement cycle $= c(t_R,d)/e(t_R,d)$. Then $e(t_R,d) = E(\min\{T_{t_R,d}, T\}) = W_{t_R}(d)$ and

$$\begin{aligned} e(t_R,d) &= K_R P(T \leq t_R) + (C+K)P(t_R < T \leq T_{t_R,d}) + CP(T_{t_R,d} < T) \\ &= K_R P(T \leq t_R) + (C+K)[P(T \leq T_{t_R,d}) - P(T \leq t_R)] + CP(T_{t_R,d} < T) \\ &= C + KP(T \leq T_{t_R,d}) - (C+K-K_R)P(T \leq t_R) \\ &= C + KQ_{t_R}(d) - (C+K-K_R)P(T \leq t_R). \end{aligned}$$

APPENDIX 4
Calculation of the cost function

To calculate the cost function $\Phi_{t_R}(d)$, we have to calculate $P(T \leq t_R)$, $Q_{t_R}(d)$ and $W_{t_R}(d)$. In the following we will suppress the subscript t_R from notation. Let $d > 0$ be the "control limit", $T_d = \max\{t_R, \min\{t\colon h(t,Z(t)) \geq d\}$ – the stopping rule, and $\tilde{T} = \min\{T, T_d\}$ – time to replacement. Define $\varphi(t,Z(t)) = h(t,Z(t))I(t \geq t_R)$. Then $T_d = \min\{t\colon \varphi(t,Z(t)) \geq d\}$. Let $\Delta > 0$ be the length of an interval selected for an approximation of the process $Z(t)$, i.e. $Z(s) = Z(i\Delta) = z_i$, $i\Delta \leq s < (i+1)\Delta$, $i = 0,1,2,\ldots$, so that the replacement decision whether $\tilde{T} > t$, $i\Delta \leq t < (i+1)\Delta$, depends on t and $z_i, z_{i-1},\ldots,z_1,z_0$. Note that whether $T > i\Delta$, depends on $z_{i-1},\ldots,z_1,z_0$. Let $p_i^d(0) = P(Z(0) = i,\ \varphi(0,i) < d) = P(Z(0) = i)I(\varphi(0,i) < d)$, and $p^d(0) = [p_i^d(0)]_i$ be a row vector. Then $P(\tilde{T} > 0,\ Z(0) = i) = P(T > 0,\ Z(0) = i,$ $\varphi(0,i) < d) = P(Z(0) = i,\ \varphi(0,i) < d) = p_i^d(0)$. Note it is possible that $\sum p_i^d(0) < 1$. Let us introduce some more notation:

$$\begin{aligned} I(l,i) &= I(\varphi(s,i) < d \text{ for all } s, l\Delta \leq s < (l+1)\Delta), \\ I(l,i,x) &= I(\varphi(s,i) < d \text{ for all } s, l\Delta \leq s \leq l\Delta + x),\ 0 \leq x < \Delta, \\ R(l,i,\Delta) &= P(T > (l+1)\Delta | T > l\Delta, z_l = i)I(l,i) \end{aligned}$$

$$\begin{aligned}
&= \exp\Big\{-\int_{l\Delta}^{(l+1)\Delta} h(s,i)ds\Big\} I(l,i),\\
R(l,i,x) &= P(T > l\Delta + x | T > l\Delta, z_l = i) I(l,i,x)\\
&= \exp\Big\{-\int_{l\Delta}^{l\Delta+x} h(s,i)ds\Big\} I(l,i,x),\ 0 \le x < \Delta,\\
D_l &= \text{diag}[R(l,i,\Delta)]_i - \text{ diagonal matrix},\\
R_l(x) &= [R(l,i,x)]_i - \text{ column vector},\\
P_l &= [p_{ij}(l)] - \text{ matrix of transition probabilities (2)}, \bar{P}_l = D_l P_l,\\
B_0 &= p^d(0),\\
B_{i+1} &= B_i \bar{P}_i.
\end{aligned}$$

(a) $P(\tilde{T} > j\Delta + x) = B_j R_j(x)$, $0 \le x < \Delta$. $P(T > j\Delta + x)$ is obtained by setting $d = \infty$, i.e. $I(l,i) = I(l,i,x) = 1$ for all l.

Proof:

Let $0 \le x < \Delta$. Then

$$\begin{aligned}
I_j(x) &= I(\varphi(s,Z(s)) < d \text{ for all } s, 0 \le s \le j\Delta + x) = \Big[\prod_{l=0}^{j-1} I(l,z_l)\Big] I(j,z_j,x), \text{and}\\
P(\tilde{T} > j\Delta + x) &= \sum_{z_0,z_1,\ldots z_j} P(T > j\Delta + x, \varphi(s,Z(s)) < d,\quad 0 \le s \le j\Delta + x, z_0, z_1, \ldots, z_j)\\
&= \sum_{z_0,z_1,\ldots,z_j} P(T > j\Delta + x, z_0, z_1, \ldots, z_j) I_j(x)\\
&= P(T > 0, \varphi(0,z_0) < d, z_0) \sum_{z_0,z_1,\ldots,z_j} \prod_{l=0}^{j-1} P(T > (l+1)\Delta | T > l\Delta, z_1)\\
&\quad \times P(z_{l+1} | T > (l+1)\Delta z_l) I(l,z_l) P(T > j\Delta + x | T > j\Delta, z_j) I(j,z_j,x)\\
&= p^d(0)\Big[\prod_{l=0}^{j-1} D_l P_l\Big] R_j(x) = B_j R_j(x), \quad j = 0,1,2,\ldots
\end{aligned}$$

(b) Let $\tau_j = \int_0^\Delta R_j(x)dx = \Big[\int_o^\Delta R(j,i,x)dx\Big]_i$, be a column vector. Then

$$W(d) = E(\tilde{T}) = \sum_{j=0}^\infty B_j \tau_j.$$

Proof:

$$\begin{aligned}
E(\tilde{T}) &= \int_0^\infty P(\tilde{T} > t)dt = \sum_{j=0}^\infty \int_{j\Delta}^{(j+1)\Delta} P(\tilde{T} > t)dt = \sum_{j=0}^\infty \int_0^\Delta P(\tilde{T} > j\Delta + x)dx\\
&= \sum_{j=0}^\infty B_j \int_0^\Delta R_j(x)dx = \sum_{j=0}^\infty B_j \tau_j.
\end{aligned}$$

Note that $A_j = B_j \tau_j \downarrow 0$, because $A_j = \int_{j\Delta}^{(j+1)\Delta} P(\tilde{T} > t)dt$, and $P(\tilde{T} > t)$ is non-increasing in t. If $\sup_i \sup\{t\colon \varphi(t,i) < d\} < \infty$, e.g. if $h(t,Z(t)) \to \infty$, $t \to \infty$, such as in the Weibull regression model for $\beta > 1$, then for some $j_0 \ge 0$, $A_j = 0$, $j \ge j_0$. The calculation can stop much earlier, when for some given ε, $r_k = \sum_{j\geq k+1} A_j \leq \varepsilon$. Let, e.g. $A_j \sim ab^{-j}$, $a > 0$, $b > 1$, then $r_k \sim \sum_{j\geq k+1} ab^{-j} \leq \ln(b)A_k \sim \ln(A_{k-1}/A_k)A_k$, and calculation can stop when $(\ln(A_{k-1}) - \ln(A_k))A_k \leq \varepsilon$.

(c) Let $F(l,i) = P(l\Delta < T \leq (l+1)\Delta,\ \varphi(s,z_l) < d$, for all $s, l\Delta \leq s \leq T | T > l\Delta,\ z_l = i)$, and $F_l = [F(l,i)]$, be a column vector. Then

$$Q(d) = P(T_d \geq T) = \sum_{j=0}^{\infty} B_j F_j.$$

Proof:

As T is continuous, and T_d is not a function of T, it is easy to prove that $P(T_d = T) = 0$. Then

$$\begin{aligned}
P(T_d \geq T) &= P(T_d > T) = \sum_{j=0}^{\infty} P(T_d > T, j\Delta < T \leq (j+1)\Delta) \\
&= \sum_{j=0}^{\infty} \sum_{z_0,\ldots,z_j} P(j\Delta < T \leq (j+1)\Delta, \varphi(s, Z(s)) < d, \quad s \leq T,\ z_0, \ldots, z_j) \\
&= \sum_{j=0}^{\infty} \sum_{z_0,\ \ldots,z_j} P(T > j\Delta, z_0, \ldots, z_j) \\
&\qquad \times I(\varphi(s, Z(s)) < d \text{ for all } s, 0 \leq s < j\Delta) F(j, z_j) \\
&= \sum_{j=0}^{\infty} B_j F_j.
\end{aligned}$$

Note that $F(l,i) = 1 - R(l,i,x(l,i))$, where $x(l,i) = \sup\{x\colon 0 \leq x \leq \Delta, \sup_{l\Delta \leq u \leq l\Delta + x} h(u,i) < d\}$.

(d) To incorporate the regular maintenance interval t_M in the calculation, we need slight changes in the calculation of B_j. Let, for simplicity, Δ is selected such that for some integer $k_0 > 1$, $t_M = k_0\Delta$, so that the maintenance is performed at points $nk_0\Delta$, $n = 1, 2, \ldots$ Then $p_{i,j_M}(nk_0 - 1) = P(z_{n,k_0} = i_M | T > nk_0, z_{n,k_0-1} = i) = 1$, i.e. $P_{n,k_0-1} = [\delta_{i_M,i}]$, so that $\bar{P}_{n,k_0-1} = [R(nk_0 - 1, i, \Delta)\delta_{i_M,i}]$.

Notes:

For the Weibull regression model, formulas for $R(l,i,\Delta)$, $R(l,i,x)$, $F(l,i)$ can be further expanded. With an obvious modifications, regarding initial conditions, the presented method can be used for the calculation of the "remaining life" characteristics, such as $E(\tilde{T}|\tilde{T} > j\Delta, z_j)$ and $P(T > t | T > j\Delta, z_j)$. The method is very convenient for the calculation, because B_j is calculated recursively. It appears from the derivation that the presented method can be applied to any stopping rule (stopping time) $\check{T}$ which is of the "multiplicative" form, i.e.

$$\begin{aligned}
I(\check{T} > j\Delta + x) &= \varphi_j(z_0, z_1, \ldots, z_j, x) \\
&= \varphi_0(z_0)\varphi_1(z_1)\ldots\varphi_{j-1}(z_{j-1})\varphi_j(z_j, x), \quad 0 \leq x < \Delta,\ j = 0, 1, 2, \ldots .
\end{aligned}$$

The calculation for the stopping rules of other types would be much less convenient in general.

REFERENCES

1. Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993), "Statistical Methods Based on Counting Processes", Springer, New York.

2. Ansell, J.I. and Phillips, M.J. (1989), "Practical Problems in the Statistical Analysis of Reliability Data", *Applied Statistics*, 38, 205-247.

3. Aven, T. and Bergman, B. (1986), "Optimal Replacement Times – a General set-up", *Journal of Applied Probability*, 23, 432-442.

4. Basawa, I.V. and Rao, P. (1980), "Statistical Inference for Stochastic Processes", Academic Press, London.

5. Cox, D.R. and Oakes, D. (1984), Analysis of Survival Data, Chapman and Hall, London.

6. Dekker, R. (1996), "Application of maintenance optimization models: A review and Analysis", *Reliability Engineering and System Safety*, 51, 229-240.

7. Dekker, R. and Scarf, P.A. (1998), "On the Impact of Optimization Models in Maintenance Decision Making: The State of Art", *Reliability Engineering and System Safety*, 60, 111-119.

8. Gill, R.D. and Johansen, S. (1990), "A survey of product-integration with a view towards application in survival analysis" *Annals of Statistics*. 18, 1501-1555.

9. Jardine, A.K.S., Makis, V., Banjevic, D. and Ennis, M. (1998), "The Optimization of Condition-based Maintenance Decisions", Proceedings of ICOMS-98, Adelaide, Australia, Paper 049.

10. Jardine, A.K.S., Ralston, P., Reid, N., Stafford, J. (1989), "Proportional Hazards Analysis of Diesel Engine Failure Data", *Quality and Reliability Engineering International*, 5, 207-216.

11. Kalbfleisch, J.D. and Prentice, R.L. (1980), "The Statistical Analysis of Failure Time Data", Wiley, New York.

12. Kebir, Y. (1991), "On Hazard Rate Processes", *Naval Research Logistic*, 38, 865-876.

13. Keiding, N. and Andersen, P.K. (1989), "Nonparametric Estimation of Transition Intensities and Transition Probabilities: a Case Study of a Two-state Markov Process", *Applied Statistics*, 38, 319-329.

14. Makis, V. (1995), "Optimal Control of a Tool-Wear Process", Proceedings of the 13th ICPR Conference, Jerusalem, 262-264.

15. Makis, V. and Jardine, A.K.S. (1991), "Optimal Replacement in the Proportional Hazards Model", *INFOR*, 30, 172-183.

16. Marzec, L. and Marzec, P. (1997), "On fitting Cox's regression model with time-dependent coefficients", *Biometrika*, 84, 901-908.

17. Murphy, S.A. and Sen, P.K. (1991), "Time-dependent coefficients in a Cox-type regression model", *Stoch. Proces. Applic.*, 39, 153-180.

18. Pitman, J.W. and Speed, T.P. (1973), "A Note on Random Times", *Stochastic Processes and their Applications*, 1, 369-374.

19. Press, W.H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. (1994), "Numerical Recipes in C", Cambridge University Press, Cambridge.

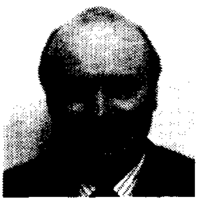
20. Scarf, P.A. (1997), "On the Application of mathematical models in maintenance", *European Journal of Operational Research*, 99, 493-506.

21. Self, S.G. and Prentice, R.L., "Commentary on Andersen and Gill's 'Cox's Regression Model for Counting Processes: A Large Sample Study' ", *Annals of Statistics*, 10, 1121-1124.

22. Tibshirani, R. (1998), "Estimating Transformations for Regression via additivity and variable stabilization", *Journal of The American Statistical Association*, 83, 394-405.

23. Toms, L.A. (1996), "Adaptive Trend Analysis – A Simple Solution to Data Variability", Proceedings of The Technology Showcase, Integrated Monitoring, Diagnostics, and Failure Prevention Conference, Mobile Alabama, University of Wales, Swansea, 48-53.

Dragan Banjevic is a Research Associate in the Department of Mechanical and Industrial Engineering and Visiting Professor in the Department of Statistics, University of Toronto. His research interests are in theoretical and applied probability, especially in reliability. His papers have appeared in *Statistics and Probability Letters, Journal of Applied Probability, Theory of Probability and its Applications* and other journals.



Andrew Jardine is a Professor in the Department of Mechanical and Industrial Engineering at the University of Toronto. His teaching and research interests lie in the general area of Engineering Management with a special interest in maintenance. He has had practical experience in devising maintenance and replacement procedures for a number of national and international organizations. He is author of *Maintenance, Replacement and Reliability* and co-editor of *Maintenance Excellence: Optimizing Life Cycle Deicisions.*

Viliam Makis is a Professor in the Department of Mechanical and Industrial Engineering, University of Toronto. His teaching and research areas include quality and reliability engineering with special interest in modeling and optimization of stochastic systems. His articles have appeared in *Mathematics of Operations Research, Technometrics, IIE Transactions, IEEE Transactions on Reliability, Journal of Applied Probability, Naval Research Logistics, EJOR, INFOR, Journal of the OR Society, International Journal of Production Economics, IMA Journal, Kybernetika,* etc. He is a Senior Member of the Institute of Industrial Engineers and of the American Society for Quality.

Marguerite Ennis obtained her B.Sc. in Statistics at the University of Stellenbosch in South Africa and her M.Sc. and Ph.D. degrees at the University of Toronto. She is interested in the application of statistics in both the medical and industrial fields and works as a freelance consultant in data analysis and statistical graphics.