

Real-time health prognosis and dynamic preventive maintenance policy for equipment under aging Markovian deterioration

ARGON CHEN* and G. S. WU

Graduate Institute of Industrial Engineering, National Taiwan University, 1,
Roosevelt Rd. Sec. 4, Taipei 106, Taiwan

(Revision received February 2006)

An often seen practice of preventive maintenance (PM) is to construct a machine's reliability model based on its historical failure records. The reliability model is then used to determine the PM schedule by minimizing the machine's long-run operation cost or average machine downtime. Machines in many hi-tech manufacturing sectors are using sophisticated sensor technologies to provide sufficient immediate online data for real-time observation of equipment condition. Not only is the historical data but also the real time condition now available for scheduling a more effective PM policy. This research is to determine an effective PM policy based on real-time observations of equipment condition. We first use the multivariate process capability index to integrate the equipment's multiple parameters into an overall equipment health index. This health index serves as the basis for real-time health prognosis under an aging Markovian deterioration model. A dynamic PM schedule is then determined based on the health prognosis.

Keywords: Real-time equipment monitoring; Equipment health prognosis; Markov chain modeling; Condition-based maintenance

1. Introduction

With the improvement of manufacturing technology and the emergence of high-tech industries, there are more and more factory productivity challenges. A major challenge is the drastically increased investment and operation cost. Take the semiconductor fabrication industry for example. The cost of a typical 25,000 wafers-per-month 300 mm fab is expected to exceed US\$2.0–3.0 billion. The huge factory investment triggers the urgent need to improve the operation effectiveness. Among various types of costs, equipment cost usually takes up the largest portion of capital investment. It is estimated that 75% of total investment will be attributed to equipment investment in a typical 300 mm fab. The utilization and effectiveness of the equipment have become extremely important for industries with intensive capital investment in equipment.

A number of Japanese industries are now implementing the concept of Total Productive Maintenance (TPM) to drive the improvement of manufacturing

*Corresponding author. Email: achen@ntu.edu.tw

efficiency (Takahasi and Osada 1990). The TPM paradigm has been shown to greatly improve the maintenance procedure, and to reduce or eliminate setups, test procedures and idle time. Similar objectives can be found in recent concerns on Overall Equipment Effectiveness (OEE) of the plant (Leachman 1997). To sum up, increasing the utilization and reducing the operation cost of equipment have become the critical factors of enhancing the company competitiveness.

To improve the effectiveness and minimize the equipment downtime, an appropriate maintenance policy is critical. On the one hand, a frequent preventive maintenance (PM) schedule reduces the system's unexpected breakdowns but also increases the maintenance cost. On the other hand, a less frequent PM schedule has less maintenance cost but increases the cost of unplanned breakdowns. How to determine an adequate policy of preventive maintenance is therefore an important issue of shop floor control.

A usual practice of PM policy in many industries is to replace parts or perform maintenance when the equipment's running time reaches a pre-determined time length (Barlow and Hunter 1960, Blanks and Tordan 1986, Elsayed 1996). The advantage of this strategy is easy to follow. Nevertheless, it depends only on the equipment's reliability model derived from historical performance instead of from the real-time condition of the equipment. Recent developments on sensors, chemical and physical nondestructive testing, and sophisticated measurement technologies have made possible real-time observations of the system performance via many kinds of on-line data (e.g. temperature, pressure, voltage, current, vibration, corrosion, fluid, etc.). Makis and Jardine (1991, 1992) proposed a condition-based maintenance (CBM) to incorporate the machine condition into the age-based reliability model. Using Cox's proportional hazards model (PHM) (e.g. Cox 1975, Cox and Oakes 1984, Kumar and Klefsjo 1994), CBM is based on a failure rate which is a function of both the tool age and the tool conditions. The same research team led by Professor Jardine has gone on to develop a decision-making software and apply CBM to different types of components or machines in various industries (see Jardine *et al.* 1997, 2001, Banjevic *et al.* 2001, Lin *et al.* 2004). The state values of the tool condition considered in CBM, however, are limited to those stochastically increasing over time and those having nondecreasing effect on the hazard rate (section 3, assumptions 2 and 4 in Makis and Jardine 1991).

There also exists a huge body of literature on optimal maintenance policies for systems under Markovian or semi-Markovian deterioration. Good reviews in this topic can be found in Barlow and Proschan (1965), McCall (1965), Sherif and Smith (1981), Valdez-Flores and Feldman (1989) and Dekker (1996). Under the deterioration models, some papers are concerned with optimal PM policies (e.g. Kao 1973); some focus on optimal inspection schedules (e.g. Milioni and Pliska 1988) and others consider simultaneous optimization of both inspection and PM schedules (e.g. Yeh 1996). Nevertheless, the optimal inspection schedule would not be a concern when modern sensor technologies are widely used especially in high-tech manufacturing industries.

To characterize a machine's health, various types of equipment data should be accounted for. With the modern sensors built in the advanced processing equipment, there are usually tens or even hundreds of data items collected. It is necessary to integrate these data items and develop a single integrated index. Not only does the index provide an easy reading for engineers to have a quick idea on the equipment's

overall performance, it also serves as a basis for determining an optimal PM policy. Gertsbakh (1977) has proposed using Discriminant Analysis to find a linear combination function of machine parameters that best distinguishes between a 'good' machine and a 'failed' one. Similar to Discriminant Analysis, a linear combination function of parameters with the maximum contribution to the machine condition can be found through Principal Component Analysis or Singular Value Decomposition (e.g. Stamatis *et al.* 1992). In this paper, it is not our objective to compare and find the best health index. Rather, we attempt to follow the modern equipment monitoring approach (Chen *et al.* 1998) to evaluate the machine condition by observing the distribution of the machine parameters' readings and comparing it against pre-defined machine specifications. A health evaluating method is thus proposed based on the ideas of multivariate process capability indices.

With a single health index available, we then attempt to propose a tool health prognosis method under an aging Markovian deterioration model of the equipment health. In using only the Markov chain to characterize a deteriorating machine, it is assumed that the transition probabilities are only state-dependent. That is, the probability to make transition to a less healthy state does not increase with the age. In the above-mentioned literature, the semi-Markovian model is often used to capture the aging effect by incorporating a random sojourn time in each state (see Kao 1973). Different from the semi-Markovian model, we introduce an aging factor that discounts the probabilities of transitions to healthier states while increasing the probabilities of transitions to less healthy states. We refer to this Markovian deterioration with aging effect as *aging Markovian* deterioration. With the equipment health prognosis, we can predict the behaviour of the equipment condition. The PM schedule with respect to either downtime or cost minimization can be then determined dynamically based on the predicted equipment health. We refer to it as *dynamic PM policy*.

As shown in figure 1, the upper dotted-line rectangle includes steps to build the offline models. Historical data are collected first. Some multivariate statistical methods, e.g. principle component analysis, and time series analysis (Chen *et al.* 1998), are applied to extract important data items in the data conversion/preprocessing step. The health index is then computed using the pre-processed equipment data. After computing the health index, we enter the second phase where an aging Markovian deterioration model is identified and estimated based on historical observations of equipment health data. By doing so, the long-run performance of the equipment is established and possibly predicted. Finally, an effective PM policy is developed accordingly. The lower dotted-line rectangle shows the online use of this proposed PM policy. The details will be stated in the rest of this paper.

2. Evaluation of equipment condition and equipment health index

The advanced equipment and sensor technologies have provided more and more immediate data to reveal a machine's condition. However, such data have not been effectively analysed and put to use in practice due to their enormous volume and the lack of efficient analysis methods. If we can establish a mechanism to analyse these real-time data, the equipment condition can be then observed and evaluated in a more timely fashion. A 'Health Index' for the equipment is therefore proposed here.

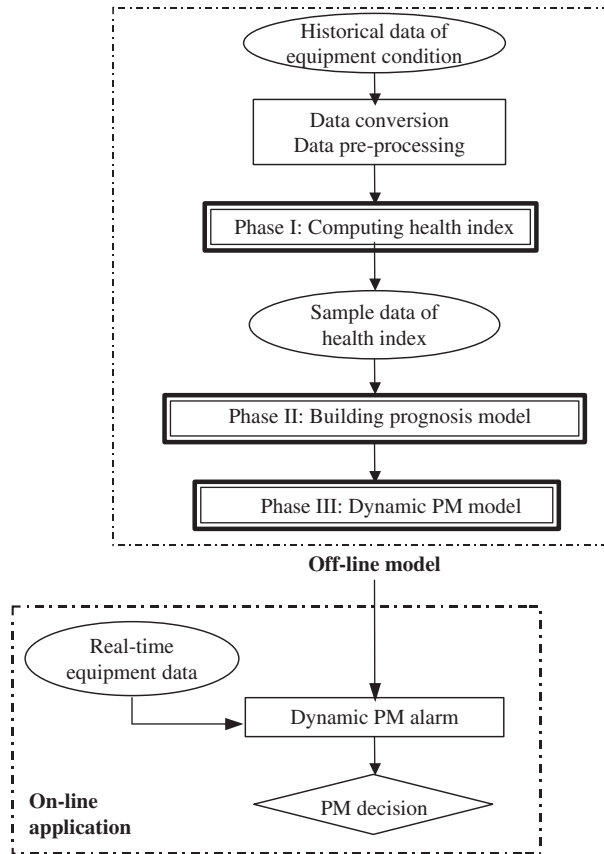


Figure 1. Dynamic PM scheme.

2.1 Evaluation of equipment condition with multivariate data

Usually, equipment provides several types of data. To integrate multiple correlated data items into a single index, multivariate statistical methods are used in this research. The health index must take into account the information contained in a set of parameters, which are online observed from the machine. For example, temperature, pressure, voltage, current, vibration, corrosion, fluid, etc., are typical parameters collected from a machine. They are often cross-correlated. Hence, not only the variation of individual variables but also the co-variation among variables should be taken into consideration.

Figure 2 shows a typical real-time equipment monitoring scheme referred to as 'Bull's Eye' scheme. Values of various machine data items are displayed simultaneously on a monitoring board. The board consists of 3 concentric circular regions with different colours: green for SAFE, yellow for WARNING, and red for DANGEROUS. The distance between an observed data point and the board's centre represents the data item's deviation from its preset target. The distribution of data points provides an easy reading of the equipment's current operating condition.



Figure 2. Equipment simultaneous monitoring scheme.

When the points are concentrated around the centre, it indicates a good overall health. In contrast, when data points are scattered over a wide area, it indicates a worrisome situation. Thus, the engineers can easily read the equipment status by examining the distribution of the data points. Mathematical treatment to create such a 'Bull's Eye' monitoring scheme will not be reiterated in this paper and can be found in O'Sullivan *et al.* (1996) and Chen *et al.* (1998).

Using the idea of 'Bull's Eye' and different from Gertsbakh's (1977) approach, an overall equipment health index can be calculated by the distribution of the data points. Using this idea, a quality measure known as multivariate process capability indices (PCIs) can be utilized. Different from our objectives here, the multivariate PCIs are created to measure a process's performance by examining resulting products' quality characteristics. But the basic theory is also feasible for evaluating equipment capability. Let's recall the *basic* idea of multivariate process capability indices as in equation (1) (Taam *et al.* 1991, Kotz and Johnson 1993).

$$C_p = \frac{\text{volume of specification region}}{\text{volume of region containing 99.73\% of observation variate}} \quad (1)$$

The basic assumption of above multivariate PCI is that the observations of machine parameters must follow multivariate normal distributions. This assumption may not be appropriate for all situations. Fortunately, many studies on data modeling, which removes the common cause or data pattern of equipment parameters, have been proposed. Most of the machine parameters can therefore be filtered with only white noise left (Chen *et al.* 1998). In fact, as long as the distribution approximately concentrates on the central region, the idea of multivariate PCI can be adopted. Based on equation (1), figure 3 shows a possible relation between the specification region and the performance region for two hypothetical quality characteristics X_1 and X_2 . The desired X_1 and X_2 are to have a target at $[0, 0]^T$, covariance equal to 0, and both variances equal to 1. The actual

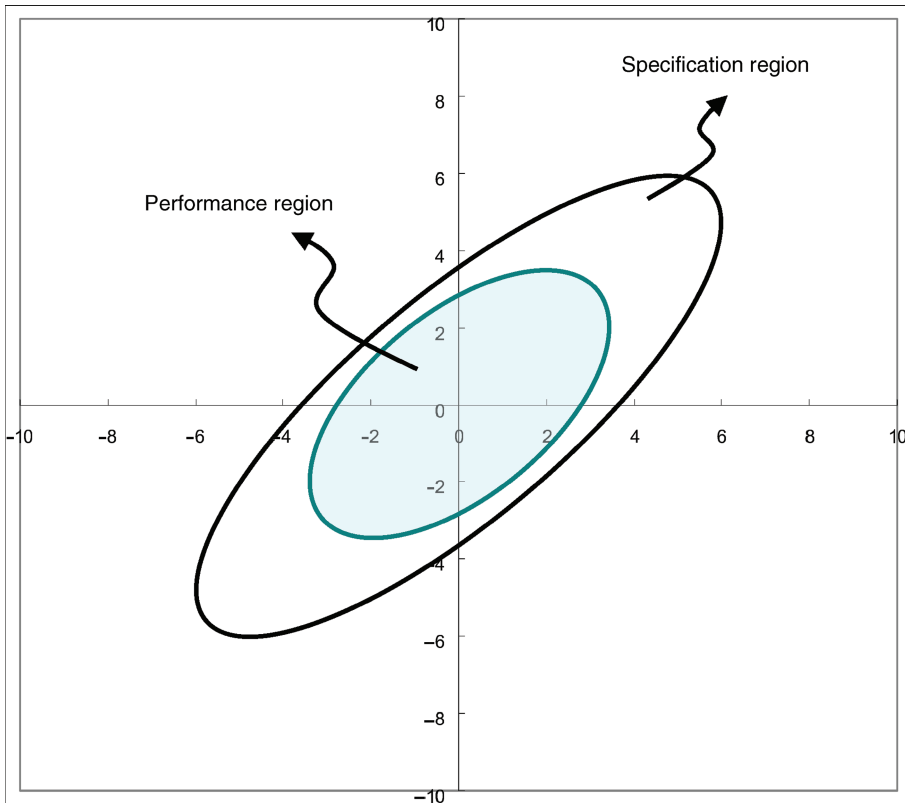


Figure 3. Performance region vs. specification region.

performance, of which 99.73% is shown in the shaded elliptical region, shows elliptical region with the covariance equals 0.6 and covers a much smaller area as compared to the specification region.

Assume that there are ν important machine parameters, X_1, X_2, \dots, X_ν . When the machine is running normally, the machine parameters will follow a multivariate normal distribution with a mean vector equal to target settings T and covariance matrix A . By replacing the product quality characteristics in the multivariate PCI with machine parameters, we define a *Machine Capability Index* (MCI) as follows

$$MCI = \frac{\text{volume of } (X - T)^T A^{-1} (X - T) \leq K^2}{\text{volume of } (X - \bar{X})^T V_0^{-1} (X - \bar{X}) \leq \chi_{\nu, 0.9973}^2} \tag{2}$$

where

- K is used to adjust the size of the specification region;
- X is the vector of sample observations of machine parameters;
- \bar{X} is the mean vector of sample observation X ;
- V_0 is the observed covariance matrix of X ;
- $V'_0 = V_0 + (\bar{X} - T)(\bar{X} - T)^T$; and

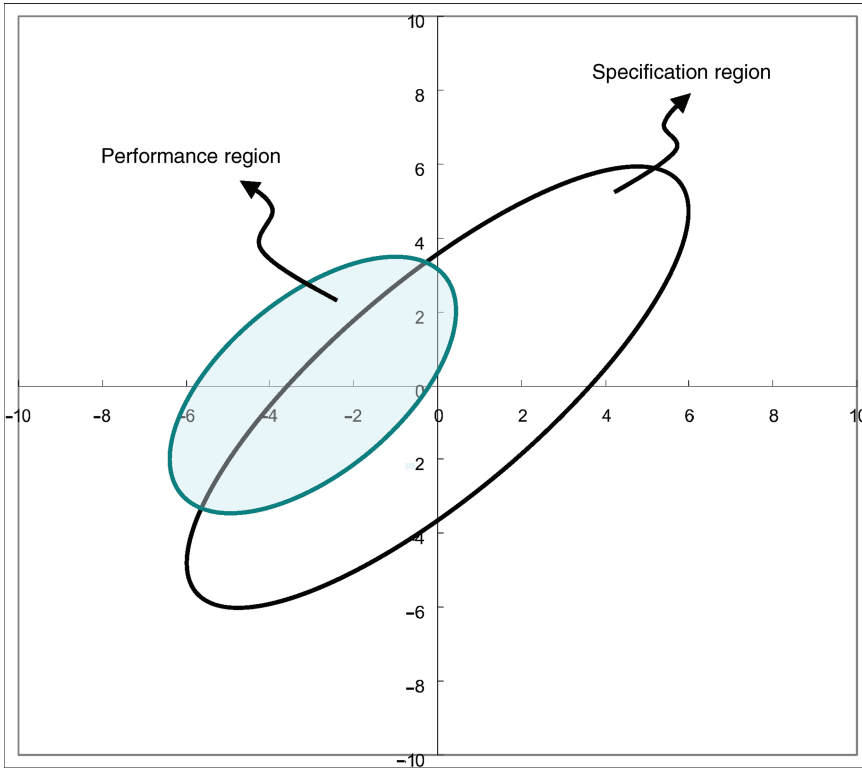


Figure 4. An off-target performance region.

$\chi^2_{\nu, 0.9973}$ is the 99.73th percentile of χ^2 distribution with ν equal to the number of machine parameters.

In figure 3, the specification region is formed by:

$$K = 6, \quad T = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \text{and} \quad A = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

while the performance region is formed by:

$$\chi^2_{2, 0.9973} = 11.83, \quad \bar{X} = T, \quad \text{and} \quad V'_0 = V_0 = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}.$$

With the above specifications and actual machine performance, the MCI in equation (2) can now be calculated as:

$$\begin{aligned} MCI &= \frac{\text{volume of } (X - T)^T A^{-1} (X - T) \leq 6^2}{\text{volume of } (X - \bar{X})^T V_0^{-1} (X - \bar{X}) \leq 11.83} = \frac{6^2 / \sqrt{\det(A^{-1})}}{11.83 / \sqrt{\det(V_0^{-1})}} \\ &= \frac{6^2 / \sqrt{2.778}}{11.83 / \sqrt{1.563}} = 2.28. \end{aligned}$$

Figure 4 shows a machine's performance region with its centre, i.e. \bar{X} , shifted from $[0, 0]^T$ to $[-3, 0]^T$.

In the case of figure 4, V'_0 is no longer the same as V_0 :

$$V'_0 = V_0 + (\bar{X} - T)(\bar{X} - T)^T = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix} + \begin{bmatrix} -3 \\ 0 \end{bmatrix} \begin{bmatrix} -3 & 0 \end{bmatrix} = \begin{bmatrix} 10 & 0.6 \\ 0.6 & 1 \end{bmatrix}.$$

The calculated MCI, thus, becomes much smaller:

$$\begin{aligned} MCI &= \frac{\text{volume of } (X - T)^T A^{-1}(X - T) \leq 6^2}{\text{volume of } (X - \bar{X})^T V_0^{-1}(X - \bar{X}) \leq 11.83} = \frac{6^2/\sqrt{\det(A^{-1})}}{11.83/\sqrt{\det(V_0^{-1})}} \\ &= \frac{6^2/\sqrt{2.778}}{11.83/\sqrt{0.1037}} = 0.588. \end{aligned}$$

It should be noted that in equation (2) specification region defined by T , A and K is assumed known in this paper. In reality, to estimate and obtain such a specification region requires prior engineering knowledge and understanding of equipment behaviour from continuous observations. Interested readers can refer to Taam *et al.* (1991) and Chen and Tsai (2004) for the details of establishing an elliptical engineering specification region.

2.2 A probable health index

One of our purposes for establishing the health index is to provide a numerical indicator of the equipment's condition. That is, we like to grade the machine's performance based on real-time evaluation of the machine condition. First, let's consider the properties of the multivariate PCI proposed earlier to evaluate the machine condition. For a process to be in a working condition, the performance region containing 99.73% of observations should fall inside the specification region. Namely, a process with the PCI value less than 1 would be deemed incapable. For the process capability to be acceptable, it is generally said that the PCI value should reach at least 1.33. Moreover, $\text{PCI} = 2$ implies that the volume of the specification region is twice as large as the volume of the performance region with the mean equal to the target. That is, the specification region contains almost all possible observations and with only a very slim chance that the equipment will perform out-of-spec. In the univariate case, when $\text{PCI} = 2$ there would be only a chance of 0.002 in one million that the equipment breaks down, i.e. out of the specification region (Kotz and Johnson 1993). Therefore, a process with $\text{PCI} \geq 2$ is said to be in an excellent condition. From a barely working condition at $\text{PCI} = 1$ to an acceptable condition at $\text{PCI} = 1.33$ and a superb condition above 2, the PCI scale, though familiar to most engineers, is ranging from 0 to ∞ and is not linearly reflecting the equipment health.

To provide a more comprehensible view on the equipment condition, the MCI proposed in the previous section can be translated into a score in a linear scale with a closed range, such as $[0, 100]$. Here we propose a possible transformation function to transform the MCI value in $(0, \infty)$ to a score in $(0, U)$ with 0 and U representing the worst and the perfect machine conditions, respectively:

$$H = \frac{1 - e^{-(MCI/d)}}{1 + e^{-(MCI-a)/d}} \times U. \quad (3)$$

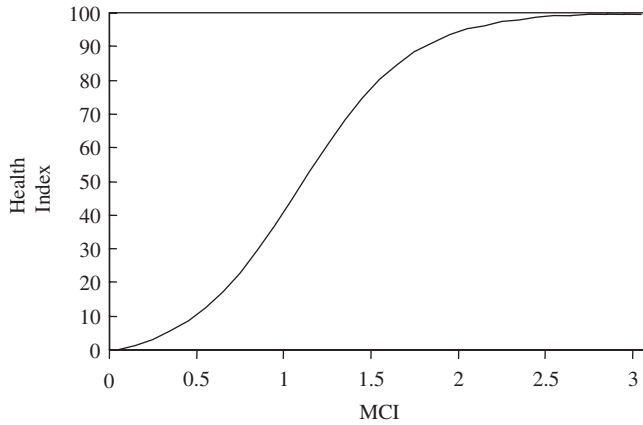


Figure 5. Mapping function— $(a, d) = (1.04, 0.32)$ for a (0, 100) score range.

Table 1. Data list of (2–8) with setting $(a, d) = (1.04, 0.32)$

MCI	Health index
0	0.00
0.5	12.34
1	44.82
1.2	60.78
1.33	70.11
1.5	80.06
2	95.07
3	99.77
4	99.99

The shape of the mapping function in equation (2) can be determined by adjusting the value of a and d based on the engineers’ own engineering judgment. a and d are parameters used to adjust the shape of the mapping function. Generally speaking, a affects the location and d affects the width and slope.

Here, we suggest a setting of $(a, d) = (1.04, 0.32)$. Figure 5 shows the curve of this mapping function with $U = 100$ representing a perfect machine condition. Table 1 lists the corresponding values of MCI and health index under this setting. A function with this setting translates $MCI = 1.33$ to about $H = 70$ and $MCI = 2$ to about $H = 95$.

It has to be noted that one may define one’s own transformation function to translate the MCI value to any closed interval as long as engineers’ view of machine condition can be faithfully reflected. Since PCI itself is already a familiar measure to most engineers, using MCI directly as a health index without any transformation is also plausible. Irrespective of the transformation methods, the health prognosis models and dynamic PM policies proposed in the coming sections can be still applied.

3. Equipment health prognosis model

In the previous section, we proposed a health index to evaluate the real-time equipment conditions. With the proposed health index, a machine with a lower score is more likely to fail than a machine with a higher score. For the purpose of determining an appropriate PM schedule, the tendency towards failure needs to be modeled. Here, we propose a failure prognosis model, which can help us to predict the machine conditions for the time to come.

3.1 Equipment health prognosis under Markovian deterioration

In practice, the equipment health based on real-time parameters variability is subject to many noises in the manufacturing environment. To make robust PM decisions based on the health index, the health index should be further classified into a few discrete states. For instance, the Bull’s eye scheme classifies the equipment health into only three states (green, yellow and red). For three equipment states, a machine with MCI above 2.0 (health score > 95), generally considered as ‘good’, can be classified as ‘Green’ while the machine with MCI value below 1.0 (health score < 45) is said to be in the ‘Red’ state (or failure state). The ‘Yellow’ state can be then defined for MCI values in between 1.0 and 2.0. It is often an engineering call to determine the number of equipment states. Although it is not the focus of this paper, the engineers should be aware that the number of states affects the sensitivity and robustness of PM decisions and the engineering judgment should be made to strike a balance.

Suppose now the health score has been classified into n discrete states: $F, 1, 2, 3, \dots, n-1$. Since the equipment health state can be evaluated at each sampling time point, it can be viewed as a stochastic process: $H = \{H_t; t \geq 0\}$. If $H_t = i$, the equipment is said to be in state i at time t . We assume here that when the process is in state i , there is a fixed probability $P_{i,j}$ that the health index will be in state j at the next time point. We characterize such a stochastic process using a *Markov chain* model. For a Markov chain, the conditional distribution of any future state H_{t+1} given the earlier states H_0, H_1, \dots, H_{t-1} becomes (Ross 2000)

$$P(H_{t+1} = j | H_t = i, H_{t-1} = i_{t-1}, \dots, H_1 = i_1, H_0 = i_0) = P(H_{t+1} = j | H_t = i) = P_{i,j}$$

We first establish a condition-based prognosis model based on Markov chain theories. Let Φ denote the matrix of one-step transition probabilities $P_{i,j}$:

$$\Phi = \begin{matrix} & \begin{matrix} F & 1 & 2 & \dots & n-1 \end{matrix} \\ \begin{matrix} F \\ 1 \\ 2 \\ \vdots \\ n-1 \end{matrix} & \begin{bmatrix} P_{F,F} & P_{F,1} & P_{F,2} & \dots & P_{F,n-1} \\ P_{1,F} & P_{1,1} & P_{1,2} & \dots & \\ \vdots & & & \ddots & \vdots \\ \vdots & & & \dots & P_{n-2,n-1} \\ P_{n-1,F} & P_{n-1,1} & P_{n-1,2} & \dots & P_{n-1,n-1} \end{bmatrix} \end{matrix} \quad (4)$$

where $\sum_{j=F}^{n-1} P_{i,j} = 1$. $P_{i,j}$ is the probability that H transits to state j given its current state is at i . If a machine breaks down, H enters the failure state F and continues to

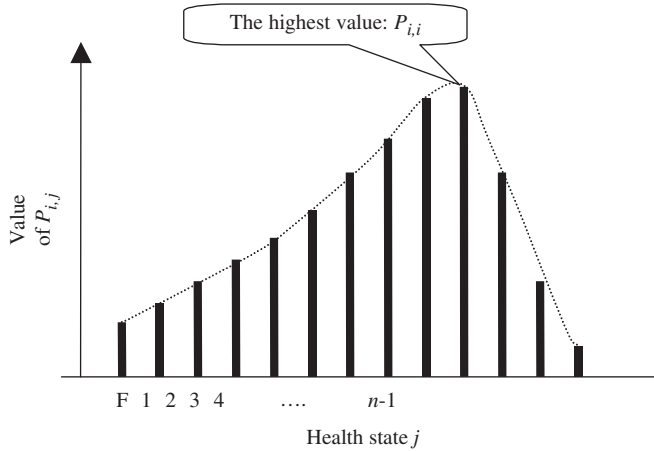


Figure 6. Bell shape of conditional state transition probability.

stay in the failure state until it is repaired. The failure state is also known as an absorbing state:

$$P_{F,F} = 1, P_{F,j} = 0, \quad \text{for } j = 1, 2, \dots, n - 1$$

The two-step transition probability can be obtained by taking the square of Φ . For instance, the (i, j) th entry of Φ^2 is the probability to be at state j after two periods of time, given the current state at i . Recursively, we can calculate the probability for the condition of the equipment to be at certain state after any number of time periods.

To further interpret the matrix Φ , each row, say row i , of Φ represents a state probability distribution given the current state at i . This conditional probability distribution will likely form a bell shape. For example, in the beginning of operation, H is most likely to be at a good-condition state, say state i , and remain at state i at the next available time point. H is less likely to move to other states. Thus, $P_{i,i}$ is likely to be the highest probability among $P_{i,j}$'s (see figure 6).

For example, let the current equipment health state be 80. Then, at the next time point the equipment health has the greatest possibility to stay at state 80. When the health state does change, the probability to move to state 70 will be greater than the probability to move to state 60. Likewise, the probability to move to state 90 should be greater than that to move to state 100. This leads to a bell-shape probability distribution as illustrated in figure 6. The above Markov-chain model of the equipment health does not explicitly consider the equipment's age although equipment aging may be exhibited through the stochastic evolution of the health state. However, with the constant Φ , when a machine arrives at a state, the transition probability distribution to move to other states is always the same no matter how old the machine has grown. That is, a machine could stochastically arrive at the same state at different ages but the deterioration 'tendency' is exactly the same given the same health state. In the following section, we will propose an aging Markovian model in which even starting with

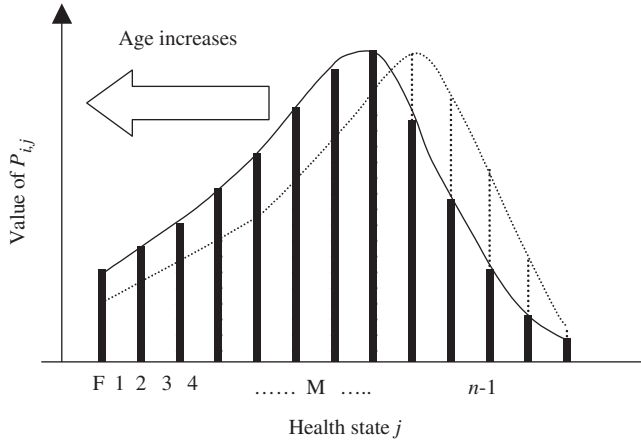


Figure 7. Wave-moving transition-probability distribution for an aging machine.

the same health state the machine becomes more likely to enter into less healthy states as it grows older.

3.2 Equipment health prognosis under aging Markovian deterioration

An important property of the above Markovian model is the bell-shape probability distribution presented by entries in the same row of Φ . But as the machine grows older the health tends to worsen due to deterioration of the equipment; i.e. the probability for the machine to become healthier will decrease while the probability of becoming less healthy will increase. Let a larger value of the state correspond to a higher health score. We can then observe that each row of matrix Φ , i.e. the conditional probability distribution, should act like a moving wave as the equipment ages. Let’s observe the wave motion in figure 7. Stochastically, a machine could visit the same state i at different ages, the machine at a higher age should have higher probabilities to become less healthy while the probabilities to become healthier should decrease. The moving-wave effect is the result of the machine’s deterioration over its run time. The longer the machine has been run, the worse the machine’s health ‘tendency’. We extend the Markovian model to include this aging effect and call it an aging Markovian deterioration model.

Suppose Φ_t is the transition probability matrix at age t and $P_{i,j}(t)$ be the entry of Φ_t representing the transition probability from state i to state j at t . Also, let $P_{iM}(t) = \max_j\{P_{ij}(t), j = F, 1, 2, \dots, n - 1\}$, where $M = \arg \max_j\{P_{i,j}(t), j = F, 1, 2, \dots, n - 1\}$, represent the peak transition probability in the bell-shape distribution. Denote the left-hand side and right-hand side cumulated probabilities as $P_i^L(t) = \sum_{j=F}^{M-1} P_{i,j}(t)$ and $P_i^R(t) = \sum_{j=M}^{n-1} P_{i,j}(t)$, respectively. Then, $P_i^R(t) = 1 - P_i^L(t)$ since $\sum_{j=1}^n P_{i,j}(t) = 1$. As shown in figure 7, when the machine becomes older, P_i^L increases while P_i^R decreases. That is, $P_i^L(t) \leq P_i^L(t + \Delta)$ and $P_i^R(t) \geq P_i^R(t + \Delta)$ where Δ is the time interval between the health observation at t and the next available observation. Assuming further that the increase

(or decrease) in $P_{i,j}(t)$ is proportional to the fraction $P_{i,j}(t)$ takes in $P_i^L(t)$ (or in $P_i^R(t)$), we have

$$P_{i,j}(t + \Delta) - P_{i,j}(t) = \frac{P_{i,j}(t)}{P_i^L(t)} [P_i^L(t + \Delta) - P_i^L(t)] \quad \text{for } j < M \quad (5a)$$

and

$$P_{i,j}(t) - P_{i,j}(t + \Delta) = \frac{P_{i,j}(t)}{P_i^R(t)} [P_i^R(t) - P_i^R(t + \Delta)] \quad \text{for } j < M. \quad (5b)$$

We define an aging factor $\delta \in (0, 1)$ as

$$\frac{P_i^L(t + \Delta) - P_i^L(t)}{P_i^L(t)} = \frac{P_i^R(t) - P_i^R(t + \Delta)}{P_i^R(t)} = \delta \quad \text{for } \forall i \text{ and } \forall t. \quad (6)$$

δ represents the increase rate in $P_i^L(t)$ for a machine becoming Δ older and remains constant for all states over the entire lifetime of an aging machine. We can now derive an age-dependent model to describe the wave-moving probability distribution from equations (5a), (5b) and (6). Given the current age t , the transition probability at the next available time $t + \Delta$ is,

$$P_{i,j}(t + \Delta) = P_{i,j}(t) + \frac{P_{i,j}(t)}{P_i^L(t)} \cdot P_i^L(t) \cdot \delta = P_{i,j}(t) \cdot (1 + \delta) \quad \text{for } j < M \quad (7a)$$

and

$$P_{i,j}(t + \Delta) = P_{i,j}(t) - \frac{P_{i,j}(t)}{P_i^R(t)} \cdot P_i^R(t) \cdot \delta = P_{i,j}(t) \cdot \left(1 - \frac{P_i^R(t)}{P_i^R(t)} \delta\right) \quad \text{for } j \geq M. \quad (7b)$$

This function is derived from the principle that $P_{i,j}$ should decrease for $j \geq M$ and increase for $j < M$. δ can be viewed as an aging factor and also should be estimated from historical equipment data. The above equations are developed to describe the wave-moving phenomenon and ensure that $\sum P_{i,j}$ remains 1. Starting with the initial transition probability matrix at $t=0$ (Φ_0), the values of the entries in Φ_t evolve according to equations (7a) and (7b) as the machine age (t) grows. The equipment health prognosis can be thus established under this aging Markovian deterioration.

3.3 Parameter estimation

Both the aging factor δ and the initial transition probability matrix Φ_0 need to be estimated. We first find a δ estimator for an aging machine given Φ . An MLE (Maximum Likelihood Estimation) method is proposed here. The problem is to find a good estimator $\hat{\delta}$ given observed values of a random sample x_1, x_2, \dots, x_n (Hogg and Tanis 1996). Now we assume there are n independent failure events that occurred at times $t_i, i=1, 2, \dots, n$. For each failure event, say the failure at t_i , there exists a stochastic process of the health index states through time t_i , denoted

by $\bar{S}(t_i)$. Let the failure occur between the k_i th and $(k_i + 1)$ th samples; i.e. $k_i\Delta \leq t_i < (k_i + 1)\Delta$ and Δ denotes the fixed time interval between two consecutive health observations. Thus, the probability that $\bar{S}(t_i)$ takes a particular path $\bar{s}(t_i)$ is

$$\begin{aligned} \Pr\left(T = t_i, \bar{S}(t_i) = \bar{s}(t_i) | \hat{\delta}\right) &= \Pr\left(k_i\Delta \leq t_i < (k_i + 1)\Delta, S_1 = s_1^{(i)}, \dots, S_{k_i} = s_{k_i}^{(i)} | \hat{\delta}\right) \\ &= \Pr\left(S_1 = s_1^{(i)}, \dots, S_{k_i} = s_{k_i}^{(i)}, S_{k_i+1} = F | \hat{\delta}\right) \\ &= \Pr\left(s_{k_i+1}^{(i)} = F | s_{k_i}^{(i)}, \hat{\delta}\right) \cdot \prod_{m=1}^{k_i} \Pr\left(s_m^{(i)} | s_{m-1}^{(i)}, \hat{\delta}\right) \end{aligned}$$

where $S_j^{(i)}, j = 1, 2, \dots, k_i$, is the j th state in the particular path $\bar{S}(t_i)$ leading to the failure state F at t_i . To rearrange it,

$$\Pr[\bar{S}(t_i) = \bar{s}(t_i) | \hat{\delta}] = \prod_{m=1}^{k_i+1} P_{s_{m-1}^{(i)}, s_m^{(i)} | \hat{\delta}} \tag{8}$$

where $P_{s_{m-1}^{(i)}, s_m^{(i)} | \hat{\delta}}$ denotes the transition probability that state changes from $s_{m-1}^{(i)}$ to $s_m^{(i)}$ given that the aging factor $\delta = \hat{\delta}$. With n failure sample paths, the likelihood function can be written as follows.

$$L = \prod_{i=1}^n \Pr(\bar{S}(t_i)) = \prod_{i=1}^n \prod_{m=1}^{k_i+1} P_{s_{m-1}^{(i)}, s_m^{(i)} | \hat{\delta}} \tag{9}$$

and the log-likelihood function is:

$$\log L = \sum_{i=1}^n \sum_{m=1}^{k_i+1} \log P_{s_{m-1}^{(i)}, s_m^{(i)} | \hat{\delta}} \tag{10}$$

Hence, the estimator of δ can be found by maximizing the log-likelihood function. That is:

$$\hat{\delta}^* = \arg \max_{\hat{\delta}} \{\log L\} = \arg \max_{\hat{\delta}} \left\{ \sum_{i=1}^n \sum_{m=1}^{k_i+1} \log P_{s_{m-1}^{(i)}, s_m^{(i)} | \hat{\delta}} \right\}. \tag{11}$$

Since $\delta \in (0, 1)$, numerical methods to search over the range $(0, 1)$ can be employed to find the optimal $\hat{\delta}^*$. In figure 8, $\log L$ as a function of $\hat{\delta}$ for $\delta = 0.002, 0.008$ and 0.01 are illustrated. Figure 8a shows the log-likelihood function for $0 \leq \hat{\delta} \leq 1$, while figure 8b shows the function in a smaller interval: $0 \leq \hat{\delta} \leq 0.04$. From figure 8, it can be conjectured, although not rigorously proven, that the log-likelihood function is a unimodal function of $\hat{\delta}$ and thus $\hat{\delta}^*$ can be found rather effectively through numerical search.

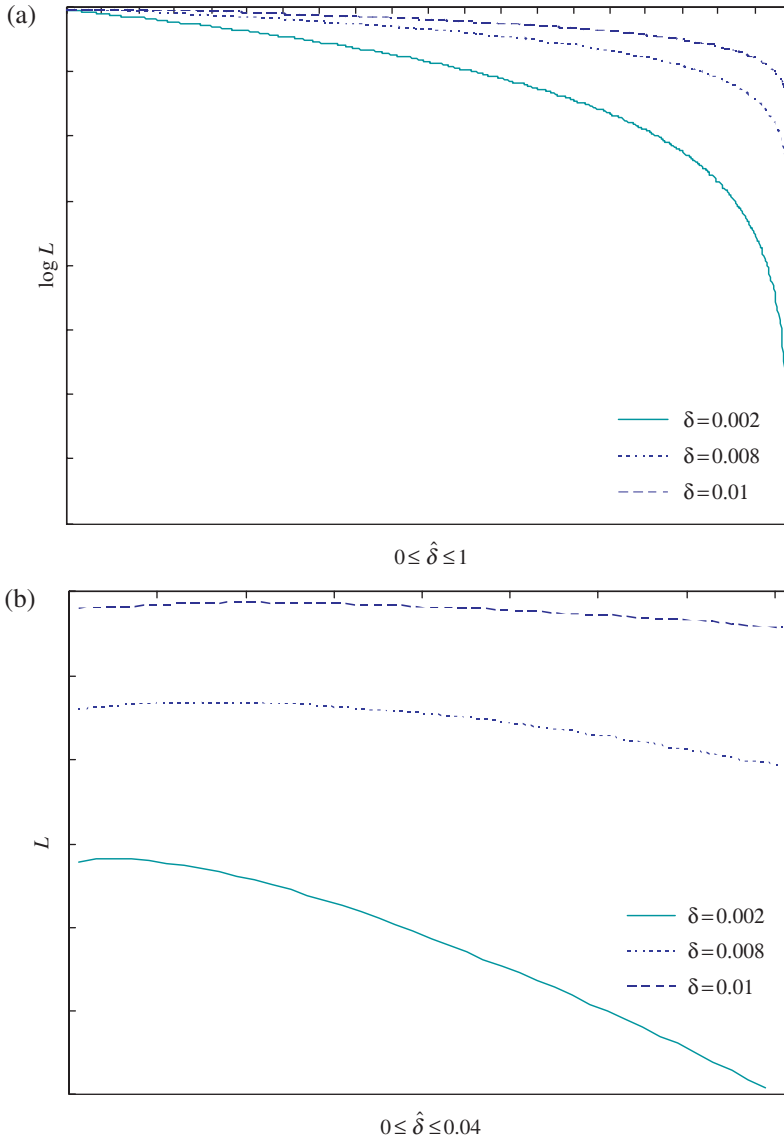


Figure 8. $\log L$ vs. $\hat{\delta}$.

3.4 Example

A numerical example is used to illustrate the construction of the aging Markovian prognosis model. Suppose $H < 70$ (or $MCI < 1.33$) is regarded as an absolutely dangerous condition for the equipment. That is, when the health index has a score of less than 70, the machine is deemed to be in a failure state. We now divide $H = [0, 100]$ into 11 states as shown in table 2.

Table 2. Health index and its corresponding states (example).

Health score	State	Health score	State
<70	Failure	85–87	6
70–72	1	88–90	7
73–75	2	91–93	8
76–78	3	94–96	9
79–81	4	97–100	10
82–84	5		

An initial transition probability matrix, Φ_0 representing the stable equipment health stochastic process, is assumed to be:

$$\Phi_0 = \begin{matrix} & F & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ \begin{matrix} F \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{matrix} & \left[\begin{array}{cccccccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.05 & 0.35 & 0.23 & 0.17 & 0.1 & 0.05 & 0.031 & 0.01 & 0.005 & 0.003 & 0.001 & \\ 0.01 & 0.1 & 0.35 & 0.2 & 0.14 & 0.1 & 0.05 & 0.032 & 0.01 & 0.005 & 0.003 & \\ 0.005 & 0.01 & 0.105 & 0.35 & 0.2 & 0.14 & 0.09 & 0.05 & 0.035 & 0.01 & 0.005 & \\ 0.003 & 0.005 & 0.01 & 0.092 & 0.35 & 0.2 & 0.14 & 0.1 & 0.05 & 0.04 & 0.01 & \\ 0.001 & 0.003 & 0.005 & 0.01 & 0.091 & 0.35 & 0.2 & 0.14 & 0.1 & 0.07 & 0.03 & \\ 5e-4 & 0.001 & 0.003 & 0.005 & 0.01 & 0.0805 & 0.35 & 0.25 & 0.17 & 0.08 & 0.05 & \\ 2e-4 & 6e-4 & 0.001 & 0.003 & 0.005 & 0.01 & 0.0802 & 0.47 & 0.25 & 0.1 & 0.08 & \\ 1e-4 & 3e-4 & 6e-4 & 0.001 & 0.003 & 0.005 & 0.01 & 0.08 & 0.6 & 0.2 & 0.1 & \\ 8e-5 & 1.2e-4 & 3e-4 & 5e-4 & 0.001 & 0.003 & 0.005 & 0.01 & 0.08 & 0.6 & 0.3 & \\ 2e-5 & 8e-5 & 1e-4 & 3e-4 & 5e-4 & 0.001 & 0.003 & 0.005 & 0.01 & 0.08 & 0.9 & \end{array} \right. \end{matrix}$$

Using this Φ_0 and $\delta = 0.005$, we simulate 20 sample paths of health index scores, each path is of size 90. Figure 9 illustrates the first 5 paths.

As we can see, the health scores are roughly decreasing over time. We use these 20 sample paths as sample data to estimate δ . To estimate the performance of this MLE estimate of the aging factor, we repeat the simulation and estimation five times to obtain an average estimate $\hat{\delta} = 0.00504$ with quadratic estimation loss = $1.064e-07$ from the simulated data. In table 3, different values of δ are used to simulate new sample paths and the same estimation method is applied to obtain the maximum likelihood estimation of δ . It can be seen that with 20 sample paths and 90 sample scores in each path, the estimate using MLE is quite robust for different values of δ (0.01–0.0005).

We now estimate the initial transition probabilities in Φ_0 . As shown below, there are $n \times n$ entries in the matrix.

$$\Phi = \begin{matrix} & F & 1 & 2 & \dots & n-1 \\ \begin{matrix} F \\ 1 \\ 2 \\ \vdots \\ n-1 \end{matrix} & \left[\begin{array}{cccccc} P_{F,F} & P_{F,1} & P_{F,2} & \dots & P_{F,n-1} \\ P_{1,F} & P_{1,1} & P_{1,2} & \dots & \\ \vdots & & & \ddots & \vdots \\ \dots & & & & P_{n-2,n-1} \\ P_{n-1,F} & P_{n-1,1} & P_{n-1,2} & \dots & P_{n-1,n-1} \end{array} \right] \end{matrix}$$

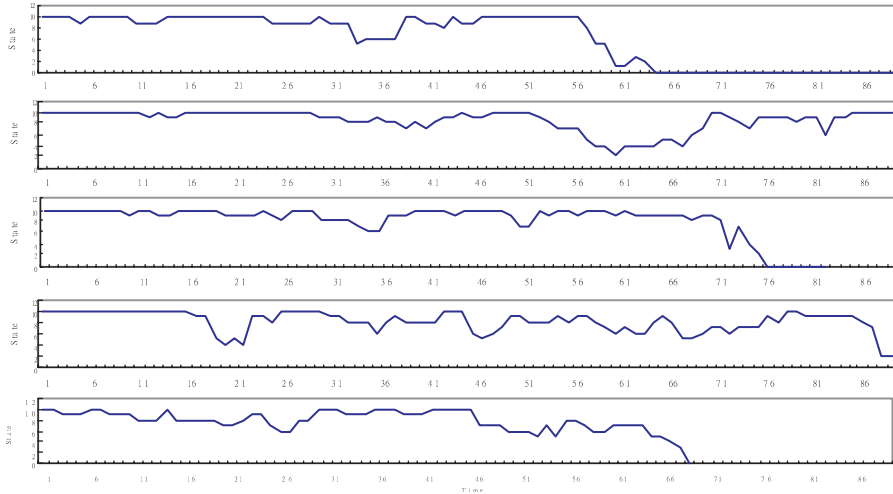


Figure 9. Five sequences of health state sample path.

Table 3. Estimation performance for different δ .

Actual δ	Average of 5 $\hat{\delta}$	Sample mean squared estimation errors
0.01	0.00946	3.54e-7
0.005	0.00504	1.06e-7
0.001	0.00108	1.06e-8
0.0005	0.00052	3.27e-9

Intuitively, it is not hard to obtain the estimators. Denote $n_{i,j}$ as the number of one-step transitions from i to j in available sample paths. The estimators of $P_{i,j}$ are given by:

$$\hat{P}_{ij} = \frac{n_{i,j}}{n_i}$$

where $n_i = \sum_{j \in S} n_{i,j}$ and is the total number of state i in sample paths. Ideally, Φ_0 should be estimated using sample paths of machines before aging, i.e. machines with the constant hazard rate during a stable-reliability period as in the well-known bath tub reliability model. However, if aging starts early, sample paths available for estimating the initial transition probabilities may be limited. Given an aging sample path and δ , we need to propose an estimate for Φ_0 . Let $m(l)$, $l=1, \dots, n_i$, be the machine ages at which the health states are i in an aging sample path. From equation (7a), we can obtain the conditional expected number of transitions from state i to state j for $j < M$:

$$\begin{aligned} & E(\text{number of transitions } i \rightarrow j | s_{m(1)} = i, s_{m(2)} = i, \dots, s_{m(n_i)} = i) \\ &= \sum_{l=1}^{n_i} P_{ij}(0)(1 + \delta)^{m(l)} \quad \text{for } j < M. \end{aligned}$$

Therefore, the initial left-hand-side transition probabilities can be estimated by:

$$\hat{P}_{ij} = \frac{n_{i,j}}{\sum_{l=1}^{n_i} (1 + \delta)^{m(l)}} \quad \text{for } j < M. \tag{12}$$

The left-hand side and right-hand side cumulated probabilities can then be estimated by:

$$\hat{P}_i^L(0) = \sum_{j=F}^{M-1} \hat{P}_{i,j}; \hat{P}_i^L(m) = \hat{P}_i^L(0)(1 + \delta)^m; \text{ and } \hat{P}_i^R(m) = 1 - \hat{P}_i^L(m) \tag{13}$$

From equation (7b), we can obtain the conditional expected number of transitions from state i to state j for $j \geq M$:

$$E(\text{number of transitions } i \rightarrow j | s_{m(1)} = i, s_{m(2)} = i, \dots, s_{m(n_i)} = i) \\ = \sum_{l=1}^{n_i} P_{ij}(0) \prod_{h=1}^{m(l)} \left[1 + \frac{P_i^L(h)}{P_i^R(h)} \cdot \delta \right] \quad \text{for } j \geq M$$

and thus we obtain the estimates for the initial right-hand side transition probabilities:

$$\hat{P}_{ij} = \frac{n_{i,j}}{\sum_{l=1}^{n_i} \prod_{h=1}^{m(l)} \left[1 + \frac{\hat{P}_i^L(h)}{\hat{P}_i^R(h)} \cdot \delta \right]} \quad \text{for } j \geq M \tag{14}$$

With equations (11)–(14), we propose the following iterative procedure to estimate δ and P_{ij} simultaneously:

- Step 1.** Give an initial guess of $\hat{\delta}$.
- Step 2.** Use equations (12)–(14) to estimate \hat{P}_{ij} given the latest $\hat{\delta}$.
- Step 3.** Use equation (11) to re-estimate $\hat{\delta}$ given \hat{P}_{ij} estimated in Step 2.
- Step 4.** Check if $\hat{\delta}$ has converged to a predetermined accuracy. If yes, stop. If not, go back to Step 2.

Since this paper’s focus is not on the estimation quality, the convergence of the above iterative procedure is not proven and should be an interesting subject for future study.

4. Dynamic preventive maintenance policy

The ultimate goal of this research is to develop a dynamic PM policy based on real-time equipment data. The main tasks of establishing a dynamic PM scheme are illustrated in figure 10. The first and second rectangles have been developed in sections 2 and 3, respectively. Our next task focuses on the third rectangle, that is, to develop an efficient dynamic PM policy. Here, we will introduce a dynamic PM scheme based on the health index and equipment’s condition prognosis model. Then, we will explain why it is an efficient method. How the dynamic PM

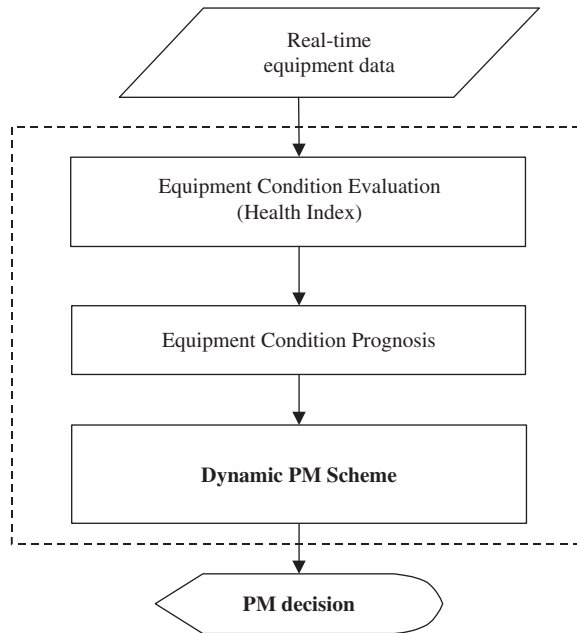


Figure 10. Overview of steps to make a dynamic PM decision.

methodology can be applied in practice will be also presented at the end of this section.

4.1 Preventive maintenance optimization models

In this paper, we consider only full-scale PMs that completely renew the equipment condition. Generally, when planning a preventive maintenance policy, the goal is to achieve two aspects of production efficiency: minimum equipment maintenance cost or minimum equipment downtime. Conventional PM models are all developed to minimize the equipment maintenance cost or downtime. Nevertheless, these conventional models (e.g. chapter 9 of Elsayed 1996) do not take into consideration the real-time condition of the equipment. PM policies are determined only based on the machine's earlier failure records. Modern sensor technology has made real-time equipment operation data available for monitoring and analysis. A more timely, adaptive PM decision can be now made in real time according to the on-line collected equipment data. This is why the word 'dynamic' is used, i.e. to perform the maintenance dynamically based on the equipment's current condition to minimize the maintenance cost or the machine's downtime ratio.

4.1.1 Cost minimization model. When minimizing the cost, we often take the long-term average of total maintenance cost as the objective cost function to be minimized. The main maintenance cost of the equipment comes from the preventive maintenance and the breakdown repair. The cost of a failure repair is usually much

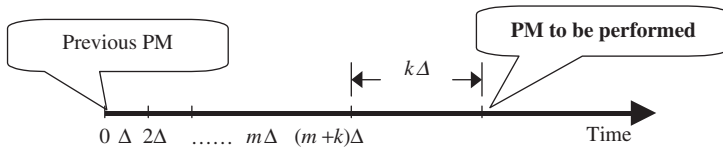


Figure 11. Illustration of PM location on time axis.

higher than that of performing a PM. Sufficient PMs can reduce the possibility of equipment breakdown but increase the PM cost. Therefore, a trade-off exists between the PM cost and the failure cost. A metric to evaluate the efficiency of a PM policy is the expected maintenance cost per unit time.

Suppose the equipment data are acquired at scheduled sampling times $\Delta, 2\Delta, 3\Delta, \dots, m\Delta, \dots$, where Δ is a constant time interval. At these specific time points, a PM decision (to perform a PM or do nothing) is made. Since the health state transition probabilities are age-dependent, Φ_{m+1} must be calculated from Φ_m using the aging Markovian model presented in section 3.2. Notice that our transition probability matrix is time dependent and is therefore a nonstationary Markov chain. That is, the equipment’s health has a tendency to worsen as it becomes older.

For a stationary Markov chain, the n -step transition probabilities $\Phi^{(n)}$ can be obtained from n powers of one-step transition probability matrix, i.e. $\Phi^{(n)} = \Phi^n$. But in the aging Markovian prognosis model, the n -step transition probabilities are different at different observation points. At any given period m , the n -step transition probabilities are calculated by

$$\Phi_m^{(n)} = \Phi_m \cdot \Phi_{m+1} \cdots \Phi_{m+n-1}. \tag{15}$$

We can now determine the expected cost per unit time, denoted by μ based on the renewal theory. If we decide to perform a PM after k periods (i.e. PM at $(m+k)\Delta$) as shown in figure 11, given a health index state s at the m th period, the expected cost per unit time is.

$$E(\mu(m, k)|s) = \frac{\text{expected cost per cycle}}{\text{expected cycle time}} = \frac{C + K \cdot \Pr\{k\Delta > T(m)|s\}}{E[m\Delta + \min(k\Delta, T(m))|s]} \tag{16}$$

where C is the cost of performing a PM, K is the additional cost for a breakdown repair, i.e. the cost of a breakdown repair is $C + K$. $T(m)$ is the Time To Failure (TTF) from the current time period m . $\Pr\{k\Delta > T(m)|s\}$ is the probability that the equipment breaks down before the PM at period k is performed given the current state is s . A maintenance cycle is terminated by either the PM or the equipment breakdown. Therefore, the numerator in equation (16) represents the expected cost per cycle and the denominator $E[\min(k\Delta, T(m))|s]$ denotes the expected cycle length. Our objective is then to find an optimal k at any given period m that minimizes $E(\mu)m, k(s)$.

In order to calculate the expected cycle length, we again use the Markov chain theory. Define the following two variables:

$f_{ij}^{(n)}(m)$: the probability that the equipment health goes from state i at period m to state j for the first time at period $m+n$

$P_{ij}^{(n)}(m)$: the probability that the equipment health state starts from i at period m and becomes j at period $m + n$

If F presents the state of machine failure and is therefore an absorbing state, it can easily be found that

$$f_{iF}^{(n)}(m) = P_{iF}^{(n)}(m) - P_{iF}^{(n-1)}(m) \quad \text{for } n > 1$$

and

$$f_{iF}^{(1)}(m) = P_{iF}^{(1)}(m) \quad \text{for } n = 1$$

where $P_{iF}^{(n)}(m)$ is an entry in $\Phi_m^{(n)}$. From the above definitions, we can immediately obtain:

$$\Pr\{k\Delta > T(m)|s\} = P_{sF}^{(k)}(m).$$

Suppose we plan a PM after k periods given the current state s at period m . $E[\min(k\Delta, T(m))|s]$ represents the expected cycle length. The cycle length is $k\Delta$ if the cycle is ended by a PM and is $T(m)$ if the cycle is ended by an equipment breakdown. The probability that the equipment is down after k periods is $P_{sF}^{(k)}(m)$, which can be obtained from $\Phi_m^{(k)}$. We can derive:

$$\begin{aligned} & E[m\Delta + \min(k\Delta, T(m))|s] \\ &= m\Delta + k\Delta \cdot \Pr\{T(m) > k\Delta|s\} \\ & \quad + \Delta \cdot \Pr\{\text{equipment fails for the first time after 1 period}|s\} + \dots \\ & \quad + k\Delta \cdot \Pr\{\text{equipment fails for the first time after } k \text{ period}|s\} \\ &= m\Delta + k\Delta \cdot (1 - P_{sF}^{(k)}(m)) + \Delta \cdot f_{sF}^{(1)}(m) + 2\Delta \cdot f_{sF}^{(2)}(m) + \dots + k\Delta \cdot f_{sF}^{(k)}(m) \\ &= m\Delta + k\Delta(1 - P_{sF}^{(k)}(m)) + \Delta P_{sF}^{(1)}(m) + 2\Delta(P_{sF}^{(2)}(m) - P_{sF}^{(1)}(m)) + \dots + k\Delta(P_{sF}^{(k)}(m) \\ & \quad - P_{sF}^{(k-1)}(m)) \\ &= m\Delta + k\Delta - \Delta \sum_{l=1}^{k-1} P_{sF}^{(l)}(m) \quad \text{for } k > 1 \end{aligned}$$

and

$$E[m\Delta + \min(k\Delta, T(m))|s] = (m + 1)\Delta \quad \text{for } k = 1. \tag{17}$$

Here, we impose that the decisions (PM, breakdown repair, or do nothing) are made and the equipment failures are observable only at discrete time points $\Delta, 2\Delta, 3\Delta, \dots, m\Delta, \dots$.

We have determined both the numerator and denominator in equation (16). Given the current state s at time $m\Delta$, if a PM is planned after k periods, the expected cost per unit time now becomes:

$$E(\mu(m, k)|s) = \frac{C + K \cdot P_{sF}^{(k)}(m)}{m\Delta + k\Delta - \Delta \cdot \sum_{l=1}^{k-1} P_{sF}^{(l)}(m)} \quad \text{for } k \geq 1 \tag{18}$$

where let $P_{sF}^{(0)}(m) = 0$. Equation (18) is then the objective cost function to be used to determine the optimal PM policy. It should be noted that a cost function, commonly used in machining economics problems (see, for example, Kaspi and Shabtay 2003), calculating the expected cost per manufactured item, instead of the expected cost per unit time, could also be formulated as the objective function. In the formulation of the expected cost per item, the maintenance cost becomes the manufacturing overhead in addition to the unit manufacturing cost. It can be easily shown that minimizing the expected cost per item is equivalent to minimizing equation (18) when the product quality cost is not considered and the processing time per item and the manufacturing cost per item are both constant.

4.1.2 Downtime minimization model. Cost estimation may be an issue for the cost minimization model. Downtime minimization model then serves as an alternative for obtaining the dynamic PM policy. There are two types of equipment downtime: downtime for performing PM and downtime for breakdown repair. The downtime for breakdown repair is usually much longer than that of performing a PM. Similar to the cost minimization, we should determine an optimal PM schedule by minimizing the long-term average downtime fraction. Let ρ denote the fraction of equipment downtime.

$$E(\rho) = \frac{\text{Total expected downtime per cycle}}{\text{Expected cycle length}} \tag{19}$$

Equation (19) expresses the expected downtime fraction. Assume R is the downtime when performing a PM and D is the additional idle time needed for a breakdown repair (i.e. the total downtime due to the machine’s breakdown is $R + D$). Similar to equation (18), the numerator in (19) can easily be found to be $R + D \cdot P_{sF}^{(k)}(m)$. The expected cycle length is, however, different from the cost model in section 4.1.1. In a cost minimization model, the downtime due to PM or breakdown repair is usually neglected.

As shown in figure 12, the PM and breakdown repair downtimes should be now taken into consideration in the expected cycle length. Suppose we are at time period m with the health state equal to s and a PM is planned to perform after k periods.

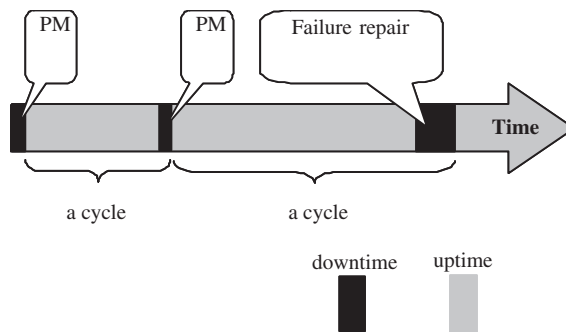


Figure 12. The cycle length for downtime minimization model.

The expected cycle length is therefore:

$$\begin{aligned}
 & E[m\Delta + \min(k\Delta + R, T(m) + R + D)|s] \\
 &= m\Delta + (k\Delta + R) \cdot \Pr\{T(m) + R + D > k\Delta + R|s\} \\
 &\quad + (\Delta + R + D) \cdot \Pr\{\text{equipment fails for the first time after 1 period}|s\} + \dots \\
 &\quad + (k\Delta + R + D) \cdot \Pr\{\text{equipment fails for the first time after } k \text{ period}|s\} \\
 &= (k\Delta + R) \cdot (1 - P_{sF}^{(k)}(m)) + (\Delta + R + D) \cdot f_{sF}^{(1)}(m) + 2\Delta \cdot f_{sF}^{(2)}(m) \\
 &\quad + \dots + (k\Delta + R + D) \cdot f_{sF}^{(k)}(m) \\
 &= (k\Delta + R)(1 - P_{sF}^{(k)}(m)) + (\Delta + R + D)P_{sF}^{(1)}(m) \\
 &\quad + (2\Delta + R + D)(P_{sF}^{(2)}(m) - P_{sF}^{(1)}(m)) + \dots + (k\Delta + R + D)(P_{sF}^{(k)}(m) - P_{sF}^{(k-1)}(m)) \\
 &= R + k\Delta - \Delta \sum_{l=1}^{k-1} P_{sF}^{(l)}(m) + DP_{sF}^{(k)}(m) \quad \text{for } k > 1
 \end{aligned}$$

and

$$E[m\Delta + \min(k\Delta + R, T(m) + R + D)|s] = (m + 1)\Delta + R + DP_{sF}(m) \quad \text{for } k = 1 \quad (20)$$

Therefore, given the current state s and the current age $m\Delta$, the expected downtime fraction is shown in equation (21) if a PM is planned after k periods.

$$E(\rho(m, k)|s) = \frac{R + D \cdot P_{sF}^{(k)}(m)}{m\Delta + R + \Delta(k - \sum_{l=0}^{k-1} P_{s,F}^{(l)}(m)) + D \cdot P_{s,F}^{(k)}(m)} \quad \text{for } k \geq 1 \quad (21)$$

In addition, the machine’s expected utilization rate can be expressed as $1 - E[\rho(m, k)|s]$.

4.2 Optimal PM decisions

The score of the health index reflects the condition of the equipment’s health. In order to determine when the machine needs a PM given a real-time score of the health index, we should find a threshold for the health index scores. When the observed score exceeds this threshold, a decision is made to perform a PM at the next available time.

Suppose that

- the current time, i.e. the machine age, is $m\Delta$
- the current state is s (i.e. the health index is corresponding to state s);
- a planned PM is schedule to perform after k periods of time (i.e. PM at time $(m + k) \Delta$).

The expected cost per unit time is $E[\mu(m, k)|s]$ and the expected downtime fraction is $E[(m, k)|s]$, which are calculated from equations (18) and (19), respectively.

Given the time point $m\Delta$ and the health index state s , there exists $k^*(s, m)$ that minimizes $E[\mu(m, k)|s]$ or $E[\rho(m, k)|s]$.

$$E(\mu(m, k^*)|s) = \min_k E(\mu(m, k)|s)$$

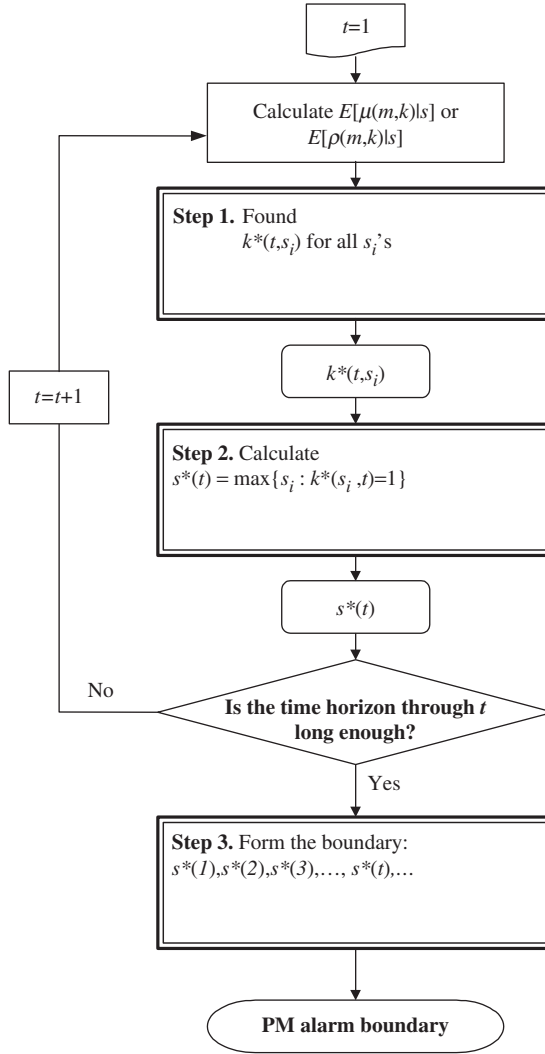


Figure 13. Steps of generating the PM alarm boundary.

or

$$E(\rho(m, k^*)|s) = \min_k E(\rho(m, k)|s) \tag{22}$$

Since we want to minimize the expected cost per unit time, a minimum cost PM decision is therefore: *if the minimum cost appears at $k^*(s, m) > 1$, it implies that to plan a PM at $k^*(s, m) > 1$ will attain a lower average cost. Then, we should not perform a PM at next decision making time. If the minimum cost appears at $k^*(s, m) = 1$, we should perform a PM right away.* This rule will be used to construct a dynamic PM policy.

If the state space is of size 11 and the state space is $S = \{F, s_1, s_2, \dots, s_{10}\}$, there will be ten k^* values, $k^*(s_1, m), k^*(s_2, m), \dots, k^*(s_{10}, m)$, at time $m\Delta$ for each S except

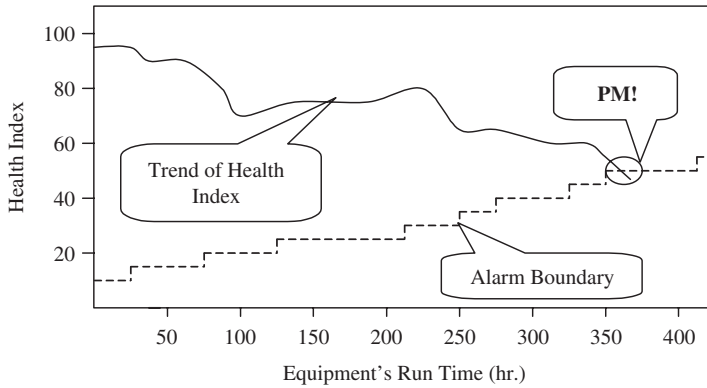


Figure 14. A typical PM alarm boundary and a sample path of health index.

for the failure state F . Suppose a larger value of S indicates a healthier equipment condition, $k^*(s_i, m)$ is then a nondecreasing function of s_i . Let

$$s^*(m) = \max\{s_i: k^*(s_i, m) = 1\} \tag{23}$$

That is, $s^*(m)$ denotes the healthiest equipment state among the equipment states that satisfy $k^*(s_i, m) = 1$ at time $m\Delta$. Then, $s^*(m)$ is the threshold for performing the PM. When a health index is calculated at time $m\Delta$, the score is first converted to a state value (e.g. table 2). The state value is then compared to the threshold state $s^*(m)$. If the observed state value is lower than $s^*(m)$ at time $m\Delta$, the decision will be to perform the PM right away.

The above decision is made at time $m\Delta$. Given any time point t , there exists a threshold state $s^*(t)$. A sequence of threshold states $s^*(t)$, $t = 1, 2, 3, \dots$, can be then calculated at time points $\Delta, 2\Delta, \dots, t\Delta, \dots$ and forms a PM alarm boundary. Figure 13 shows the steps in a flow chart.

The PM alarm boundary is actually the least tolerable value of the equipment health. That is, any observed health state lower than this boundary will be regarded as an alarming situation where a PM is urgently needed.

Figure 14 shows a typical PM alarm boundary in dotted line and a sample path of actual health state observations in solid line. As can be seen, dynamic PM decisions can be easily made by monitoring the trend of the actual health index observations. Once the trend goes below the alarm boundary, a PM should be performed.

4.3 Example

In this section, we use an example to explain how to calculate the alarm boundary from the equipment condition prognosis model. The prognosis model with an aging factor $\delta = 0.025$ presented in the example of section 3 is used. Take the downtime minimization model for example. We can calculate the expected downtime fraction for different sets of values of k , m , and s using equation (21).

Table 4 lists the expected values of downtime fractions (%) ($R = 10$ and $D = 100$) for $t = 51$ and 52 . For $t = 51$ in table 4, the minimum expected downtime calculated

Table 4. Expected downtime fraction (%) and PM threshold states.

Time	K	$s_{10}=10$	$s_9=9$	$s_8=8$	$s_7=7$	$s_6=6$	$s_5=5$	$s_4=4$	$s_3=3$	$s_2=2$	$s_1=1$	Threshold state $s^*(m)$
$t=51$	1	9.579	7.719	7.319	4.149	1.428	1.273	1.204	1.162	1.137	1.121	1.121
	2	2.308	1.540	1.356	1.237	1.243	1.256	1.279	1.317	1.373	1.454	1.454
	3	1.349	1.274	1.255	1.272	1.342	1.399	1.467	1.570	1.702	1.864	1.864
	4	1.247	1.268	1.294	1.358	1.453	1.536	1.632	1.756	1.887	2.015	2.015
	5	1.243	1.298	1.350	1.433	1.523	1.597	1.677	1.762	1.833	1.888	1.888
	6	1.257	1.322	1.381	1.457	1.516	1.559	1.601	1.636	1.657	1.669	1.669
	Min. of $E(\rho)$	1.243	1.268	1.255	1.237	1.243	1.256	1.204	1.162	1.137	1.121	4
$t=52$	1	9.568	7.565	7.149	3.913	1.391	1.260	1.197	1.158	1.135	1.120	1.120
	2	2.236	1.494	1.333	1.227	1.243	1.257	1.280	1.319	1.376	1.457	1.457
	3	1.329	1.267	1.251	1.272	1.347	1.406	1.475	1.582	1.716	1.881	1.881
	4	1.242	1.269	1.298	1.366	1.464	1.549	1.648	1.774	1.905	2.033	2.033
	5	1.245	1.305	1.360	1.447	1.538	1.612	1.693	1.776	1.845	1.897	1.897
	6	1.264	1.334	1.396	1.473	1.529	1.571	1.611	1.643	1.663	1.673	1.673
	Min. of $E(\rho)$	1.242	1.267	1.251	1.227	1.243	1.257	1.197	1.158	1.135	1.120	4

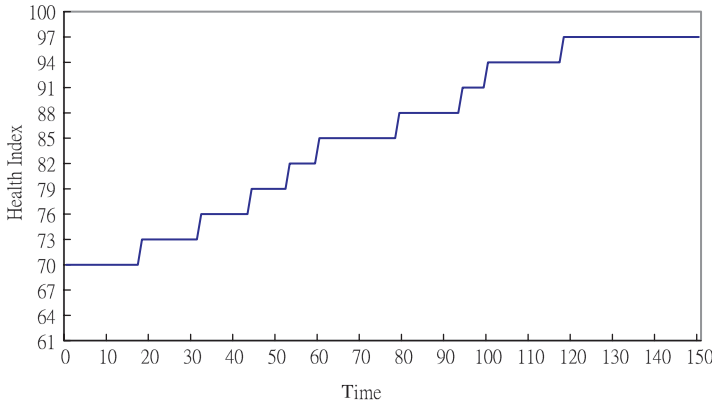


Figure 15. PM alarm boundary (example).

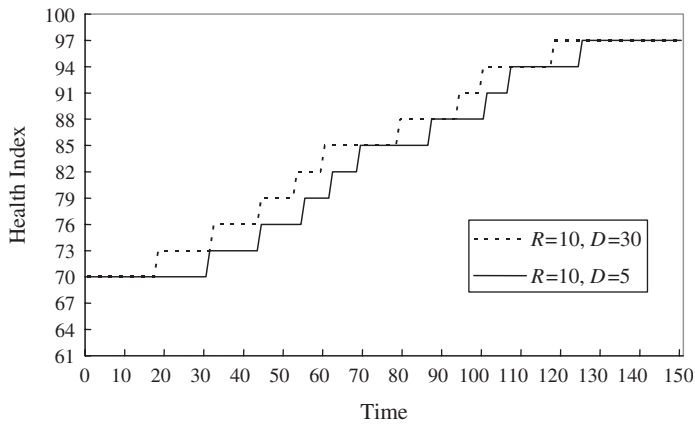


Figure 16. Comparison of PM alarm boundaries with different R and D (example).

by equation (22) is underlined for each state s_i . As can be seen, the minimum expected downtime with $k^* = 1$ appears in state 1, 2, 3, and 4. State 4 has a largest health state value so that the threshold at time $t = 51$ is set at state 4 by equation (23). Other threshold states at different time points can be obtained in the same way. The minimum downtime corresponding to the threshold states are highlighted in table 4.

Figure 15 presents the PM alarm boundary formed by the threshold states over the time horizon from 0 to 150, where the states have been translated back to the corresponding value of health index (refer to table 2).

We can find that the PM alarm boundary in figure 15 increases along with equipment's run time. This increasing boundary results from the increasing probability for the equipment's condition to become worse. When the failure probability grows up with time, the expected downtime fraction due to failure repair will increase. Therefore, the increasing rate of the PM alarm boundary is greatly affected by the relation between PM downtime and the failure repair downtime. That is, the boundary will increase rapidly if the failure repair downtime is much more than the PM down time (i.e. D is large compared to R). On the other hand, the

boundary will grow slowly with time if the additional downtime due to failure repair is relatively small. Figure 16 shows the PM alarm boundaries with different values of (R, D) . The dotted line represents the boundary with a large D and the solid line represents the boundary with a relatively small D . As shown, a smaller D will result in a lower boundary.

5. Concluding remarks

In this paper, we have demonstrated a procedure to evaluate and prognosticate the equipment health. Although the proposed health index is rather *ad hoc* and more rigorous comparisons and discussions are needed, it serves as a starting point for utilizing the vast equipment sensor data to gain information on the equipment health. This paper has also presented a novel condition-based PM policy. To distinguish our approach from the conventional condition-based maintenance (CBM) and PM policies under Markovian or semi-Markovian deterioration model in the literature, we have named our approach a dynamic PM policy under aging Markovian deterioration. In CBM, the time-based maintenance, with the failure rate directly impacted by a hazard rate function, is extended to incorporate the equipment condition as a factor precipitating the hazard rate. The dynamic PM policy, in contrast, is rooted in a stochastic health model and the age is modeled as a factor hastening the equipment health deterioration. To ensure the applicability of the proposed approach, model estimation is also developed and the procedure is demonstrated through examples. In addition, the PM policies considered in this paper are only full-scale PM policies. Optimization models for other types of PM policies, such as minimum repairs, can be studied in the future research.

Acknowledgements

The authors would like to thank Professor Gertsbakh and two anonymous referees for their valuable suggestions. Professor R.-S. Guo's skillful project management has helped the smooth completion of this research. This research is funded by NSC88-2212-E-002-063 project.

References

- Banjevic, D., Jardine, A.K.S., Makis, V. and Ennis, E., A control-limit policy and software for condition-based maintenance optimization. *INFOR*, 2001, **39**, 32–50.
- Barlow, R.E. and Hunter, L., Optimum preventive maintenance policies. *Oper. Res.*, 1960, **8**, 90–100.
- Barlow, R.E. and Proschan, F., *Mathematical Theory of Reliability*, 1965 (John Wiley: New York).
- Blanks, H.S. and Tordan, M.J., Optimum replacement of deteriorating and inadequate equipment. *Qual. Reliab. Engng Int.*, 1986, **2**, 183–197.
- Chen, A., Guo, R.-S., Yang, A. and Tseng, C.-L., An integrated approach to semiconductor equipment monitoring. *J. Chin. Soc. Mech. Enging*, 1998, **19**(6), 581–591.
- Chen, A. and Tsai, C.C., Accomodating engineering knowledge in T^2 control chart construction for equipment FDC, in *Proceedings International Symposium on Semiconductor Manufacturing*, 2004, pp. 281–285 (Semiconductor Portal: Tokyo).

- Cox, D.R., Partial likelihood. *Biometrika*, 1975, **62**, 269–276.
- Cox, D.R. and Oakes, D., *Analysis of Survival Data*, 1984 (Chapman and Hall: London).
- Dekker, R., Applications of maintenance optimization models: a review and analysis. *Reliab. Engng Syst. Safe.*, 1996, **51**, 229–240.
- Elsayed, E.A., *Reliability Engineering*, pp. 527–540, 1996 (Addison Wesley: New York).
- Gertsbakh, I.B., *Models of Preventive Maintenance*, pp. 229–244, 1977 (North-Holland: Amsterdam).
- Hogg, R.V. and Tanis, E.A., *Probability and Statistical Inference*, 4th ed., 1996 (Prentice Hall: Upper Saddle River, NJ).
- Jardine, A.K.S., *Maintenance, Replacement, and Reliability*, 1973 (Pitman/Wiley: New York). Reprinted edition available from International Academic Services, PO Box 2, Kingston, Ontario, Canada, K7L4V6.
- Jardine, A.K.S., Banjevic, D. and Makis, V., Optimal replacement policy and the structure of software for condition-based maintenance. *J. Qual. Maint. Engng*, 1997, **3**(2), 109–119.
- Jardine, A.K.S., Banjevic, D., Wiseman, M., Buck, S. and Joseph, T., Optimizing a mine haul truck wheel motors' condition monitoring program—use of proportional hazards modeling. *J. Qual. Maint. Engng*, 2001, **7**(4), 286–300.
- Kao, E.P.C., Optimal replacement rules when changes of state are semi-Markovian. *Oper. Res.*, 1973, **21**, 1231–1249.
- Kaspi, M. and Shabtay, D., Optimization of the machine economics problem for a multistage transfer machine under failure, opportunistic and integrated replacement strategies. *Int. J. Prod. Res.*, 2003, **41**(10), 2229–2247.
- Kotz, S. and Johnson, N.L., *Process Capability Indices*, pp. 37–115, 1993 (Chapman & Hall: London).
- Kumar, D. and Klefsjo, B., Proportional hazards model: a review. *Reliab. Engng Syst. Safe.*, 1994, **44**, 177–188.
- Leachman, R.C., Closed-loop measurement of equipment efficiency and equipment capacity. *IEEE Trans. Semicond. Manuf.*, 1997, **10**(1), 84–97.
- Lin, D., Wiseman, M., Banjevic, D. and Jardine, A.K.S., An approach to signal processing and condition-based maintenance for gearboxes subject to tooth failure. *Mech. Syst. Signal Proc.*, 2004, **18**, 993–1007.
- Makis, V. and Jardine, A.K.S., Optimal replacement in the proportional hazards model. *INFOR*, 1991, **30**(1), 172–183.
- Makis, V. and Jardine, A.K.S., Computation of optimal policies in replacement models. *J. Math. Appl. Bus. Indust.*, 1992, **3**, 169–175.
- McCall, J.J., Maintenance policies for stochastically failing equipment: a survey. *Manag. Sci.*, 1965, **11**, 493–524.
- Milioni, A.Z. and Pliska, S.R., Optimal inspection under semi-Markovian deterioration: basic results. *Nav. Res. Log.*, 1988, **35**, 373–392.
- O'Sullivan, P., Martinez, J., Durham, J. and Felker, S., Advanced fault detection for the semiconductor industry. *Future Fab Int.*, 1996, **1**, 71–73.
- Ross, S.M., *Introduction to Probability Models*, 7th ed., 2000 (Academic Press: San Diego).
- Sherif, Y.S. and Smith, M.L., Optimal maintenance models for systems subject to failure—a review. *Nav. Res. Log. Quar.*, 1981, **28**, 47–74.
- Stamatis, A., Mathioudakis, K. and Papaïliou, K., Optimal measurement and health index selection for gas turbine performance status and fault diagnosis. *J. Engng Gas Turb. Power*, 1992, **114**, 209–216.
- Taam, W., Subbaiah, P. and Liddy, J.W., A note on multivariate capability indices. Working paper, Department of Mathematical Sciences, Oakland University, Rochester, MI, 1991.
- Takahasi, Y., and Takashi, O., TPM: total productive maintenance. *Asian Prod. Ass.*, Tokyo, Japan, 1990.
- Valdez-Flores, C. and Feldman, R.M., A survey of preventive maintenance models for stochastically deteriorating single-unit systems. *Nav. Res. Log.*, 1989, **36**, 419–446.
- Yeh, R.H., Optimal inspection and replacement policies for multi-state deteriorating systems. *Eur. J. Oper. Res.*, 1996, **96**, 248–259.