A Comparison of Hidden Markov and Semi-Markov Modeling for a Deterioration System subject to Vibration Monitoring

Chen Lin and Viliam Makis*
Department of Mechanical and Industrial Engineering, University of Toronto,
5 King’s College Road, Toronto, ON M5S 3G8, CANADA

(Received on August 29, 2014, revised on November 12, 2014 and February 06, 2015)

Abstract: We compare a hidden Markov and Erlang semi-Markov modeling of a partially observable deteriorating system operating under a varying load and subject to multi-sensor vibration monitoring. The evolution of the unknown state process is described by a hidden, two state semi-Markov process with an Erlang sojourn time distribution in the healthy state. The unknown model parameters are estimated using the EM algorithm. We derive explicit formulae for the parameter re-estimation in the EM algorithm, which leads to a fast estimation procedure. An optimal Bayesian maintenance policy is developed minimizing the long run expected average cost per unit time and a formula for the mean residual time in the healthy state is derived as a function of the posterior probability statistic. The results show that a simple hidden Erlang model performs better than a hidden Markov model, which is encouraging for considering a more general hidden semi-Markov modeling in future research.

Keywords: Partially observable deteriorating system, condition-based maintenance, EM algorithm, hidden semi-Markov model, Erlang distribution, multivariate Bayesian maintenance control

1. Introduction

Condition based maintenance (CBM) is a maintenance program that recommends maintenance actions based on one or more indicators showing that equipment is going to fail or that equipment performance is deteriorating. The actions are taken only when there is evidence of abnormal behavior of a physical asset. This is very useful for systems such as aircraft, power plant and military equipment, to guarantee high reliability of their performance. For years, vibration condition monitoring has been proven to be an effective approach to identify deterioration of machines for CBM, fault detection, diagnostics and prognostics. In practice, the maintenance decisions need to be made according to criteria such as risk, cost, reliability and availability. Since the cost criterion applies to most situations, the development of effective maintenance policies based on cost minimization dominates in the literature.

Vibration data obtained from condition monitoring can be used for extracting useful information to determine the system condition. Due to the advancements in data measurement and computer technology, automated data collection from multiple sensors has become common in recent years. A detailed review can be found in [1]. Surprisingly, the development of cost effective CBM policies using multivariate vibration data collected under varying load is very limited in the literature.

A high level of vibration is usually caused by a hidden defect. Observed vibration data and the hidden system deterioration can be modeled using a hidden Markov model (HMM) (see e.g., [2], [3, 4]) or a hidden semi-Markov model (HSM). In practice, machines are rarely allowed to run to failure and data are commonly suspended. Vibration
data with failure information are scarce in real applications or non-existent at all, mainly because such systems are preventively maintained before failure occurs (see e.g., [5], [6]). For such critical systems without historical failure data, it is suitable to model the degradation process as a two state HSMM.

Hidden semi-Markov chain possesses the flexibility of the hidden Markov chain without the restriction of exponential or geometric distribution of the sojourn times in its hidden states. A detailed review of the HMM can be found in [7]. The phase-type modeling is a method of converting a non-Markovian model to a Markov model by enlarging the state space and this method has a wide applicability in queuing networks modeling and in many other areas. A phase-type random variable, such as simple Erlang-$m$ variable can be represented as $m$ exponentially distributed phases connected in series.

An N-state HSMM-based diagnostics and prognostics methods with univariate observation processes have been considered by some researchers ([8], [9], [10]), but the closed form expressions for maximum likelihood estimation do not exist for such models. This makes model fitting computationally very demanding.

In this paper, we focus on the development of an estimation and optimal maintenance control scheme in a two-state hidden semi-Markov model framework considering Erlang distribution for the system sojourn time in the healthy state and multivariate vibration data collected under varying load at equidistant time epochs. Observed vibration data carries partial information about the hidden system state. Data pre-processing and time series modeling are required before we can proceed to the estimation of the unknown models parameters using the so-called reference model approach [11]. One should first fit for example a vector autoregressive (VAR) model that accounts for both cross and autocorrelation in the multivariate vibration data histories when the system is in the healthy state. The residuals, which represent the observation process are then obtained using the fitted model for complete data histories. Because complete information is not available, the Expectation - Maximization (EM) algorithm is suitable to obtain the joint state and observation parameter estimates. In this paper, we derive explicit formulae for the parameter re-estimation in the EM algorithm, which leads to a fast estimation procedure.

Once the parameters of the HSMM are estimated, a procedure for determining an optimal maintenance policy based on a multivariate Bayesian control chart will be developed. Multivariate Bayesian control chart has been proved to be an optimal tool in multivariate quality control ([12], [13]). Two criteria have been widely considered for maintenance optimization over a long time horizon: availability maximization and cost minimization. Since the cost criterion applies to most situations, we consider developing a Bayesian cost-optimal CBM policy which can be used also as an early fault detection scheme based on multivariate vibration measurements collected under varying load. The procedure developed in this paper covers data pre-processing, modeling, parameter estimation and Bayesian control, which has not been considered in the maintenance control or vibration modeling before.

In condition based maintenance, the development of the optimal maintenance policy as well as the prognostic procedures based mainly on the conditional reliability function and the remaining useful life (RUL) estimation are the main objectives. A useful estimate of the RUL is the mean residual life, which is defined as the mean length of time from the current instant to the end of the system’s useful life. It is an essential reliability characteristic in health management of a gearbox [14]. Si et al. [5] commented on the methods for obtaining the remaining useful life (RUL) for two categories of models, namely, models based on directly observable processes and partially observable processes.
The conclusion was that RUL based on a model with a partially observable process is harder to achieve. Our paper presents a new approach of using a posterior probability statistic in a continuous time hidden semi-Markov framework to predict mean remaining time in the healthy state for a gearbox subject to vibration monitoring. In summary, this paper presents several new developments: a complete procedure which includes data pre-processing, modeling, parameter estimation and Bayesian control, as well as prediction of the remaining time in the healthy state using vector vibration data collected under varying load. The procedure is illustrated using real vibration data.

In Section 2, we formulate the condition-based maintenance model. In Section 3, we present a parameter estimation procedure for this model. Using the EM algorithm, both the state and observation model parameters can be estimated simultaneously. Section 4 presents a computational algorithm in the SMDP framework which can be used to obtain the optimal maintenance policy minimizing the long-run expected average cost per unit time. In Section 5, a formula for the estimation of the remaining time in the healthy state is developed. In Section 6, a case study using real vibration data is developed to demonstrate the whole procedure. The numerical results are compared with the ones obtained from a hidden Markov modelling using the same vibration data set. Section 7 provides concluding remarks.

2. Model Description

We assume that the deteriorating system follows a continuous time semi-Markov process \( \{X_t: (t \in \mathbb{R}_+)\} \) with state space \([0, 1]\). States 0 and 1 are unobservable, representing healthy and unhealthy states, respectively. The system starts in a good state, and the sojourn time \( \tau \) in this state has an Erlang distribution with parameters \( S = \{m, r\}, \) i.e., the density function is given by:

\[
g(y|x) = \frac{1}{(r/m)^m} \frac{(r/m)^{(y-x)}}{\Gamma(m)} e^{-(r/m)(x-y)}
\]

where \( \Gamma(m) \) is the gamma function.
where an integer \( m > 0 \) is the shape parameter and \( r > 0 \) is the scale parameter. The unhealthy state is absorbing.

The system is monitored at equidistant sampling epochs \((h, \ldots , Nh)\) and the vibration data \( y_n \) is obtained at the epoch \( nh \). For given data history \( i, i = 1, \ldots , M \), \( Y^i = \{ y^i_1, \ldots , y^i_{Nh} \} \) represents the collection of all \( d \) dimensional vector observation data up to time \( Nh \), which is the last sampling point for history \( i \). We assume that the observations \( \{ y^j_1 \} \) are conditionally independent and have a multivariate normal distribution given the state of the system. This assumption is valid once the appropriate data pre-processing method has been applied (see [11]), and \( \{ Y^j_i \} \) represent the residuals for history \( i \). For the remainder of the paper, the residuals are then chosen as the “observation” process satisfying the assumption of multivariate normality and conditional independence. Thus, \( y_n^i \) conditional on \( X_{nh} = x \), \( x = \{0, 1\} \), has \( d \) variate normal distribution \( N_d(\mu_x, \Sigma_x) \) with density function:

\[
g(y|x) = \exp\left(-\frac{1}{2}(y - \mu_x)'\Sigma_x^{-1}(y - \mu_x)\right).\]  

After collecting a new observation vector, a decision is made whether to run the system until the next sampling point, or stop the system and possibly carry out full preventive maintenance after an inspection, which brings the system to the healthy state. The following cost structure is considered: \( C_{OM}, C_L, C_{PM}, C_{LP}, C_s \), where the cost components are described in the Nomenclature section.

The objective is to determine the optimal maintenance policy minimizing the long-run expected average cost per unit time. From renewal theory, the expected average cost of the system is equal to the expected cycle cost divided by the expected cycle length:

\[
E(CC) = \inf_{\eta \in [0,1]} \frac{E_P(\ eta(\eta))}{E_P(\eta)}.
\]  

Assume that \( \Pi_n = P(X_n = 1|y_1, \ldots , y_n) \) represents the probability that the process is in an unhealthy state given the observations up to time \( nh \). \( E_P \) is the conditional expectation given \( P \). \( CC \) is the total cost over one complete cycle of length \( CL \). In this paper, we consider the multivariate Bayesian control approach to find the optimal maintenance policy for a deterioration system subject to multivariate vibration monitoring under varying load.

3. Parameter Estimation using Erlang Sojourn Time Distribution

Erlang distribution is a phase-type distribution obtained as a convolution of exponential distributions. Most of the original applications of this kind of distribution were in the area of queuing theory and modelling of communication systems. The cumulative distribution function of Erlang\((m, r)\) is given by:

\[
F_S(t) = \begin{cases} 
\frac{r^m t^{m-1} e^{-rt}}{(m-1)!}, & \text{if } t \geq 0 \\
0, & \text{if } t < 0
\end{cases}.
\]  

The state parameters are \( S = \{m, r\} \). Under the Erlang distribution assumption, the hidden stochastic process \( \{X_t\} \) has \( m+1 \) states \( \{1, \ldots , m, m+1\} \). The new or renewed system will go
through phases $j = \{1, \cdots, m\}$, which represent the system’s healthy state. The system will be in phase $\{m + 1\}$ if it enters the unhealthy state.

Suppose we have obtained $M$ observation histories of the form $\bar{y}^i = \{y^i_1, \cdots, y^i_{N_i}\}$, where $i = \{1, \cdots, M\}$ and $N_i$ is the length of the $i^{th}$ history. Let $\mathcal{Y} = \{\bar{y}^1, \cdots, \bar{y}^M\}$ represent all observable data and $L(O, S | \mathcal{Y})$ be the associated likelihood function, where $\{O, S\}$ are the sets of unknown observation parameters $O = \{\mu_0, \mu_1, \Sigma_0, \Sigma_1\}$ and state parameters $S = \{m, r\}$. Because the sample path $(X_t; t \in \mathbb{R}_+)$ of the deterioration process is not observable, maximizing $L(O, S | \mathcal{Y})$ analytically is not possible. The EM algorithm resolves this difficulty by iteratively maximizing the so-called pseudo-likelihood function.

A detailed review can be found in [15]. Let $\{O_0, S_0\}$ be the set of some initial values of the unknown parameters. The EM algorithm works as follows:

**E-step.** For $j \geq 0$, compute the pseudo log-likelihood function defined by:

$$
\Omega(O,S|O_j,S_j) = E_{(O_j,S_j)}(\ln L(O,S|\mathcal{Y}^c))
$$

(5)

where $\mathcal{Y}^c$ represents the complete observation data set, consisting of the observation histories augmented with the sample path information of the state process $X$.

**M-step.** Maximize the expectation computed in the first step:

$$(O_{j+1}, S_{j+1}) \in \arg \max_{O,S} \Omega(O,S|O_j,S_j).$$

(6)

$$(O_{j+1}, S_{j+1})$$ are chosen as the updated parameter estimates for the next iteration. The E and M steps are repeated until the Euclidean norm $\| (O_{j+1}, S_{j+1}) - (O_{j}, S_{j}) \| < \epsilon$ with selected small $\epsilon > 0$. $(O_{j+1}, S_{j+1})$ are then chosen as the optimal parameters $(O^*, S^*)$.

Next, we derive explicit formulae for both the likelihood function $L(O, S | \mathcal{Y}^c)$ and pseudo likelihood function $\Omega(O, S | \mathcal{Y}^c)$. The parameter estimations of Theorem 1, Lemma 1 and Theorem 2 are presented in Appendix.

4. Cost-optimal Bayesian CBM Policy

Recently, Kim et al. [16] and Jiang et al. [17] considered effective 3 state HMM-based maintenance control models based on a Bayesian control chart which monitors the posterior probability that system is in the warning state. In real situations, the sojourn time distribution can be non-exponential. In this section, we consider the Erlang healthy state sojourn time distribution in a 2-state HSMM model framework. Note that for the Erlang distribution, the sojourn time in the healthy state can be represented as a sum of $m$ independent exponential random variables, where $m$ is a positive integer. Each phase has a common exponential distribution with mean $\frac{1}{r}$. Let $\{X_t, t \geq 0\}$ be a Markov process with state space $\{1, \cdots, m, m + 1\}$, where states $\{1, \cdots, m\}$ represent the healthy state, and $\{m + 1\}$ represents the unhealthy state of the degradation process. The posterior probability that the system is in the unhealthy state becomes the probability that the system is in phase $\{m + 1\}$, which we write as $\Pi_{n+1|m+1}$. We also need to define $\Pi_{n+1,j|j} \in \{1, \cdots, m\}$, to represent the posterior probabilities that the system is in the phase $j$ at time $(n+1)h$, given all the observations until the $(n+1)^{th}$ sampling epoch. Denoting $\Pi_n = (\Pi_{n,1}, \cdots, \Pi_{n,m+1})$, the vector $\Pi_n$ becomes a multidimensional statistic, where the component $\Pi_{n,m+1}$ is used for the optimal decision making. Thus, we define the posterior probabilities as follows:

$$
\Pi_{n+1,m+1} = P(X_{n+1} = m + 1 | y_1, \cdots, y_n, y_{n+1} = y, \Pi_n),
$$

$$
\Pi_{n+1,j} = P(X_{n+1} = j | y_1, \cdots, y_n, y_{n+1} = y, \Pi_n), \; j \in \{1, \cdots, m\}.
$$

(7)
\[ \Pi_{n+1,m+1} = \frac{f(y|\mu_{n+1}) \cdot D_{m+1}}{\int f(y|\mu_{n+1}) \sum_{j \in \{1, \ldots, m\}} D_j \cdot D_{m+1} \, dy}, \quad j \in \{1, \ldots, m\}, \]  
\[ \Pi_{n+1,j} = \frac{f(y|\mu_{n+1}) \cdot D_j}{\int f(y|\mu_{n+1}) \sum_{j \in \{1, \ldots, m\}} D_j \cdot D_{m+1} \, dy}, \quad j \in \{1, \ldots, m\}, \]  
\[ D_j = \mathbb{P}(X_{n+1} = j|Y, \ldots, Y, \Pi_n) = \sum_{k \in \{1, \ldots, m\}} \mathbb{P}(X_{n+1} = j|X_n = k) \cdot \Pi_{n,k}, \quad j \in \{1, \ldots, m\}, \]  
\[ D_{m+1} = \sum_{k \in \{1, \ldots, m+1\}} \mathbb{P}(X_{n+1} = m+1|X_n = k) \cdot \Pi_{n,k}. \]

From Eq. (4), the transition probabilities in Eqs. (14-15) are as follows:
\[ \mathbb{P}(X_{n+1} = m+1|X_n = m+1) = 1, \]  
\[ \mathbb{P}(X_{n+1} = j|X_n = k) = \frac{(r_m)^{m-j} \cdot e^{-r_m}}{(j-k)!}, \quad m \geq j \geq k \geq 1, \]  
\[ \mathbb{P}(X_{n+1} = j|X_n = k) = e^{-r_k}, \quad m \geq j \geq k \geq 1, \]  
\[ \mathbb{P}(X_{n+1} = m+1|X_n = k) = 1 - \sum_{j, k=1, j \neq k} \mathbb{P}(X_{n+1} = j|X_n = k) - \mathbb{P}(X_{n+1} = k|X_n = k), \quad m \geq k \geq 1. \]  

The initial posterior probabilities are:
\[ \Pi_{0,m+1} = 1, \Pi_{0,j} = 0, \quad m \geq j \geq 1. \]  

Generally, the computation of the long-run expected average cost requires discretization of \([0,1]\), the range of the posterior probability component. The set of the SMMDP states is defined as \( S = \{0, \Pi_{n,m+1}|[0,1], PM\} \). We denote \( \Pi^* \in [0,1] \) as the optimal control limit for \( \Pi_{n,m+1} \) that minimizes the long-run expected average cost per unit time. When implementing the Bayesian control policy, the posterior probability \( \Pi_{n,m+1} \in [0,1] \) is used to monitor the deterioration process. The whole posterior probability vector must be updated after each observation. When \( \Pi^* \leq \Pi_{n,m+1} \), the system is stopped and full inspection is performed with inspection cost \( C_i \). The system has probability \( 1 - \Pi_{n,m+1} \) to be in the healthy state and probability \( \Pi_{n,m+1} \) to be in the warning state. If the system is in the warning state, preventive maintenance is performed at cost \( C_{PM} \), which takes \( T_{PM} \) time units. The associated lost production cost is incurred at a rate of \( C_{LP} \). After inspection or preventive maintenance, the system is renewed and a new cycle begins. The optimal average cost \( g(\Pi^*) \) can be obtained by defining and solving a semi-Markov decision problem. The posterior probability \( \Pi_{n,m+1} \) is plotted on the Bayesian control chart. If the current value \( \Pi_{n,m+1} \) lies in the interval \( \left[ \frac{t_{m+1}}{I}, \frac{t_{m+1}+0.5}{I} \right] \), we assume that \( \Pi_{n,m+1} = \frac{t_{m+1}+0.5}{I} \). If \( \Pi^* > \Pi_{n,m+1} \), we continue. The long-run expected average cost \( g(\Pi) \) can then be obtained by solving the following system of linear equations [18]:
\[ v_i = c_i - g(\Pi) \cdot t_i + \sum_{j \in \mathbb{S}} P_{ij} \cdot v_j, \]  
\[ v_0 = 0. \]  

The quantities such as the costs, expected sojourn times and transition probabilities can be obtained as follows:
\[ c_i = \int_0^\infty C_{OM} \cdot \mathbb{P}(X_{nh+t} = m+1|\Pi_n) \, ds + C_v, \quad \text{if} \quad \Pi_{n,m+1} = \frac{t_{m+1}+0.5}{I} < \Pi^*, \]  
\[ c_i = C_i \cdot T_i + C_{LP} \cdot T_p, \quad \text{if} \quad \Pi_{n,m+1} = \frac{t_{m+1}+0.5}{I} > \Pi^*, \]  
\[ c_{PM} = C_M \cdot T_M + C_{LP} \cdot T_{PM}. \]  

where
\[ \mathbb{P}(X_{nh+t} = m+1|\Pi_n) = \sum_{k \geq 1} P(X_{nh+t} = m+1|X_n = k, \Pi_n) \Pi_{n,k}. \]

The associated expected sojourn times \( t_i \) are given by:
The associated transition probabilities $P_{ij}$ are given by:

$$P_{ij} = P(j \leq \Pi_{n+1,m+1} \leq j^* | Y_1, \ldots, Y_m, \Pi_{n,1}, \Pi_n),$$

$$= F(j^* - 1) - F(j^*), \text{ for } \Pi_{n,m+1} < \Pi^*, \text{ where } j^* = \frac{j-1}{T} \text{ and } j = \frac{j}{T}$$

The distribution function $F(x)$ of $\Pi_{n+1,m+1}$ has the following form:

$$F(x) = Pr(\Pi_{n+1,m+1} \leq x | Y_1, \ldots, Y_m, \Pi_n),$$

$$= \sum_{j=1}^{\infty} Pr(V(y) > k(x) | X_{(n+1)h}) = 0 \cdot \frac{\beta}{\sum_{j=1}^{\infty} \Delta \Pi_{n+1} = 1} + Pr(V(y) > k(x) | X_{(n+1)h} = 1) \cdot \frac{\beta}{\sum_{j=1}^{\infty} \Delta \Pi_{n+1} = 1},$$

$$\text{and, } P_{00} = 1 - \Pi_{n+1,1} \Pi_{n+1,m+1} > \Pi^*, P_{1,0} = \Pi_{n+1,1} \Pi_{n+1,m+1} > \Pi^*, P_{y,0} = 1.$$ (17)

$F(x)$ can be computed using equations in [19], which provide a closed-form expression for the cumulative distribution function of $V(y)$.

5. Mean Remaining Time in the Healthy State

We can assess the mean remaining time of the system in the healthy state by using the posterior probability $\Pi_n$. Using Eqs. (17), the conditional reliability function considering an Erlang-2 distribution is given by:

$$R(t | \Pi_n) = Pr(\xi > nt + t | Y_1, \ldots, Y_m, \Pi_n),$$

$$= Pr(X_{n+1} = 1 | X_n = 1) + Pr(X_{n+1} = 2 | X_n = 1) \Pi_{n,1} + Pr(X_{n+1} = 2 | X_n = 2) \Pi_{n,2},$$

$$= \left( e^{-rt} + e^{-rt} \right) \Pi_{n,1} + e^{-rt} \Pi_{n,2}.$$ (18)

The mean remaining time in the healthy state, $\xi$’, using an Erlang-2 sojourn time distribution can be computed as follows:

$$\mu_{nh} = E(\xi' | Y_1, \ldots, Y_m, \Pi_n) = \int_0^{\infty} R(t | \Pi_n) dt = \frac{1}{r} \Pi_{n,1} + \frac{1}{r} \Pi_{n,2}.$$ (19)

The mean remaining time in the healthy state using an Erlang-1 distribution, saying a HMM, can be simplified as $\mu_{nh} = \frac{1}{r} (1 - \Pi_n)$. Next, we apply the estimation and optimal control methods developed in this paper to real vibration data.

6. Case Study

6.1 Experimental Data Collection

We consider the determination of the cost-optimal CBM policy and the early fault detection scheme under varying load. The vibration data considered in this section was obtained from the Mechanical Diagnostic Test Bed (MDTB) [20] built by Pennsylvania State University, Applied Research Laboratory in the Condition-Based Maintenance Department (Figure 1). This MDTB gearbox contained a 70-tooth driven helical gear and a 21-tooth pinion gear. It was driven at a set input speed using a 22.38 kW, 1750 rpm drive motor. The torque was applied by a 5.95 kW absorption motor. The test-run 14 was selected for this study. The gearbox was run at 100% output torque for 95 hours at constant load (with input speed 1750 rpm and output torque 555in-lbs), then increased to 300% torque in additional 19.3 hours with varying load (with input speed 1750 rpm and output torque 1665in-lbs). The gearbox was run under varying load from files 194 to 338, which are shown in Figure 2(a). The sampling interval between two sampling epochs is 0.1333 hours. Each sampling file is
One triaxial accelerometer set was used and selected in this experiment. The sensor A10 is in the axial direction, the sensor A11 is in the parallel to the floor direction and the sensor A12 is in the perpendicular to the floor direction.

![Testing Bed and Selected Triaxial Sensors](image)

Due to its complicated structure, vibration data pre-processing and modeling are required before applying the HSMM methodology. By observing the signals from the speed sensor (Figure 2a), the motor velocity fluctuation range was less than 0.06%, which can be considered as a constant speed. We consider time synchronous averaging (TSA) method for the vibration data pre-processing. Files 194 to 246 are used to represent the healthy condition of the gearbox. According to the changing load conditions (Figure 2b), we divide the training files 194 to 246 into four groups: 300%, (200%, 250%), (100%, 150%) and 50%. Then we consider four stationary time series processes using these four groups, which cover six load levels mentioned above. We studied four time series models using files 194-246 to represent the vibration data under varying load. The loads are stable at each sampling epoch. Four VAR models have been built for four stationary time series processes when data was sampled at six load levels: 300%, (200%, 250%), (100%, 150%), and 50%. Residuals for the complete data histories for files 247 to 338 were computed and they are shown in Figure 3(b). This study only considers the monitoring of the gearbox health condition under varying load when the system is in the early stage of deterioration, i.e., it considers files from 194 to 295 obtained from the 13.58 hours of the data collection period. The healthy portion of the observation data history is obtained by considering the data when the gearbox system operates in a healthy state. Time series models are fitted to this data which can be used as a reference model for the calculation of the residuals for the complete data histories. The orders of the VAR models are determined using both the AIC and BIC information criteria (see Table 1). This computation was conducted using statistical software R2.14.1. Multivariate Ljung-Box tests applied to AIC selected VAR model gave better results than for the models with BIC selected order. Therefore, the fitted models are selected by AIC with lags: 291, 152, 131, and 210, and the residuals are obtained in Figure 3b.

Once the fitted model residuals are obtained, some statistical indicators such as root mean square (RMS), kurtosis, crest factor, etc., can be selected in further fault diagnosis. We investigated RMS values obtained from VAR model residuals. Figure 4 shows RMS values of one deterioration history. We divided residual RMS values of files 247 to 295 into two groups. In Figure 4, we assume that from the files 247 to 283, the system is in state 0. The files 284 to 295 represent the system in state 1. The independence test and
normality test are conducted for both healthy state and warning state residual data (Table 2).

![Data modeling, 194-248, Data testing, 247-338](image)

**Figure 2:** (a) The Output Torque V05 and the Average Value of the Drive Motor Speed V01, and (b) Four Time Series Models

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>291</td>
<td>69</td>
</tr>
<tr>
<td>Model 2</td>
<td>152</td>
<td>69</td>
</tr>
<tr>
<td>Model 3</td>
<td>131</td>
<td>54</td>
</tr>
<tr>
<td>Model 4</td>
<td>210</td>
<td>67</td>
</tr>
</tbody>
</table>

**Table 1:** Vector Autoregressive Models selected with AIC and BIC Criteria

![Original signals: files 194-338](image)

(a) Original signals: files 194-338

![Time wave residuals for testing files 247-338](image)

(b) Time wave residuals for testing files 247-338

**Figure 3:** Data Preprocessing

<table>
<thead>
<tr>
<th>RMS of residuals file</th>
<th>Healthy portion (files 247 to 283)</th>
<th>Unhealthy portion (files 284 to 295)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence (Portmanteau)</td>
<td>0.1255</td>
<td>0.3656</td>
</tr>
<tr>
<td>Multivariate Normality (Henze-Zirkler)</td>
<td>0.2743</td>
<td>0.4506</td>
</tr>
</tbody>
</table>

**Table 2:** p-values of the Independence and Normality Tests
6.2 Parameter Estimation

Seventy gear teeth deterioration histories are obtained from model residuals. We choose feasible initial values and run the EM algorithm using the convergence criterion $\left| \left( O_{t+1}, S_{t+1} \right) - \left( O_{t}, S_{t} \right) \right| < 10^{-6}$. The final estimates of the model parameters using an Erlang sojourn time distribution are given in Table 3. A comparison with a hidden Markov model is shown in Table 4. All computations were coded in Matlab (2009) on an Inter Corei5, 2.6GHz with 8G RAM. It has been found that both procedures converge rapidly in less than nine seconds. Due to a more complex model, computations in a HSMM using an Erlang distribution take longer time than using a HMM. From Table 3 and Table 4, it can be seen that the mean time in a healthy state using an Erlang-2 distribution is 4.587 hours. The mean time in a healthy state using an exponential distribution is 4.34 hours.

<table>
<thead>
<tr>
<th>Table 3: Iterations of the EM Algorithm using an Erlang-2 Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial value</td>
</tr>
<tr>
<td>$\hat{P}$</td>
</tr>
<tr>
<td>$\hat{\beta}_0$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\Sigma_0 * 10^2$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\Sigma_1 * 10^2$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\Omega$</td>
</tr>
<tr>
<td>$T_{sec}$</td>
</tr>
</tbody>
</table>

6.3 The Optimal CBM Policy

Next, we compute the optimal control limit $\Pi^*$ using both Erlang distribution and exponential distribution by setting the times $T_{PM} = 10$ hours, $T_I = 1$ hour, and the costs
particular phase. The red dots data (Figure 5). Each line represents the posterior probability that the system is in that plot three posterior probabilities based on the parameter estimations from this vibration The control limits and optimal costs compared with a HMM are given in Table 7. We also Table 7 that the optimal limit equals 0.3 using the exponential distribution. Once the negative [19]. We show the application procedure by considering an Erlang-2 sojourn
eigenvalues in both healthy state and unhealthy state using Provost’s method are all negative [19]. The control limits and optimal costs compared with a HMM are given in Table 7. We also plot three posterior probabilities based on the parameter estimations from this vibration data (Figure 5). Each line represents the posterior probability that the system is in that particular phase. The red dots represent the posterior probability that the system is in an unhealthy state. At each observation epoch, the sum of all posterior probabilities equals to one.

**Table 4: Iterations of the EM Algorithm using Exponential Sojourn Time Distribution**

<table>
<thead>
<tr>
<th></th>
<th>Initial value</th>
<th>Update 1</th>
<th>Update 2</th>
<th>Update 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\vartheta$</td>
<td>0.5</td>
<td>0.2295</td>
<td>0.2003</td>
<td>0.23</td>
</tr>
<tr>
<td>$\hat{\mu}_0$</td>
<td>[1]</td>
<td>[47.2891]</td>
<td>[47.9704]</td>
<td>[47.3575]</td>
</tr>
<tr>
<td></td>
<td>36.5589</td>
<td>36.6605</td>
<td>35.4012</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19.6862</td>
<td>19.7645</td>
<td>19.4490</td>
<td></td>
</tr>
<tr>
<td>$\hat{\mu}_1$</td>
<td>[2]</td>
<td>[53.3294]</td>
<td>[62.7064]</td>
<td>[65.7481]</td>
</tr>
<tr>
<td></td>
<td>38.1972</td>
<td>41.0312</td>
<td>45.9170</td>
<td></td>
</tr>
<tr>
<td></td>
<td>21.4618</td>
<td>24.6428</td>
<td>26.0633</td>
<td></td>
</tr>
<tr>
<td>$\Sigma_0 \times 10^2$</td>
<td>0.01 0.01 0.01</td>
<td>23.0 16.7 9</td>
<td>1.19 0.25 0.24</td>
<td>0.99 0.1 0.18</td>
</tr>
<tr>
<td></td>
<td>0.01 0.02 0.01</td>
<td>16.7 13.3 6.7</td>
<td>0.25 0.62 0.08</td>
<td>0.1 0.42 0.02</td>
</tr>
<tr>
<td></td>
<td>0.01 0.03 0.03</td>
<td>9 6.7 3.8</td>
<td>0.24 0.08 0.21</td>
<td>0.18 0.02 0.17</td>
</tr>
<tr>
<td>$\Sigma_1 \times 10^2$</td>
<td>0.04 0.02 0.02</td>
<td>29 19 10.5</td>
<td>5.85 0.82 1.16</td>
<td>4.98 0.12 0.75</td>
</tr>
<tr>
<td></td>
<td>0.02 0.05 0.02</td>
<td>19 13.9 7.2</td>
<td>0.82 1.52 0.37</td>
<td>0.12 1.54 0.13</td>
</tr>
<tr>
<td></td>
<td>0.02 0.06 0.01</td>
<td>10.5 7.2 4.2</td>
<td>1.16 0.37 1.07</td>
<td>0.75 0.13 0.97</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>-2.8483e4</td>
<td>-0.6</td>
<td>-0.135</td>
<td>-5.63e-6</td>
</tr>
<tr>
<td>$T_{\text{sec}}$</td>
<td>0</td>
<td>1.7472</td>
<td>2.808</td>
<td>4.9452</td>
</tr>
</tbody>
</table>

**Table 5: Expected Average Costs and Optimal Control Limit using Erlang-2 Distribution**

<table>
<thead>
<tr>
<th>$\Pi_{n,1}$</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>67.52</td>
<td>67.53</td>
<td>67.53</td>
<td>67.92</td>
<td>68.18</td>
<td>68.22</td>
<td>68.26</td>
<td>68.30</td>
<td>70.28</td>
</tr>
<tr>
<td>0.2</td>
<td>67.21</td>
<td>67.20</td>
<td>67.19</td>
<td>67.53</td>
<td>67.56</td>
<td>67.64</td>
<td>67.64</td>
<td>67.65</td>
<td>69.32</td>
</tr>
<tr>
<td>0.3</td>
<td>66.87</td>
<td>66.85</td>
<td>66.83</td>
<td>66.70</td>
<td>67.33</td>
<td>67.33</td>
<td>67.33</td>
<td>67.30</td>
<td>68.79</td>
</tr>
<tr>
<td>0.4</td>
<td>66.76</td>
<td>66.73</td>
<td>66.71</td>
<td>66.93</td>
<td>67.06</td>
<td>67.15</td>
<td>67.13</td>
<td>67.10</td>
<td>68.49</td>
</tr>
<tr>
<td>0.5</td>
<td>66.65</td>
<td>66.62</td>
<td>66.59</td>
<td>66.78</td>
<td>66.94</td>
<td>67.03</td>
<td>67.01</td>
<td>66.98</td>
<td>68.30</td>
</tr>
<tr>
<td>0.6</td>
<td>66.61</td>
<td>66.58</td>
<td>66.55</td>
<td>66.72</td>
<td>66.86</td>
<td>66.96</td>
<td>66.93</td>
<td>66.90</td>
<td>68.17</td>
</tr>
<tr>
<td>0.7*</td>
<td>66.59</td>
<td>66.57</td>
<td>66.54</td>
<td>66.71</td>
<td>66.84</td>
<td>66.93</td>
<td>66.90</td>
<td>66.88</td>
<td>68.13</td>
</tr>
<tr>
<td>0.8</td>
<td>66.64</td>
<td>66.61</td>
<td>66.59</td>
<td>66.73</td>
<td>66.88</td>
<td>66.97</td>
<td>66.94</td>
<td>66.91</td>
<td>68.19</td>
</tr>
<tr>
<td>0.9</td>
<td>66.74</td>
<td>66.76</td>
<td>66.76</td>
<td>66.90</td>
<td>67.01</td>
<td>67.07</td>
<td>67.09</td>
<td>67.03</td>
<td>68.28</td>
</tr>
</tbody>
</table>

The optimal limit equals 0.7 using the Erlang-2 distribution (Table 6). It can be seen in Table 7 that the optimal limit equals 0.3 using the exponential distribution. Once the posterior probability that the system is in the unhealthy state is greater than $\Pi'$, full system inspection is initiated. It is noted that the expected average costs obtained using the Erlang sojourn time distribution are lower than the costs obtained using the
exponential distribution. This indicates that inspections occur more frequently in HMM than in HSMM. For a typical suspension history in Figure 4, full system inspection occurs at the data file 283 using the exponential sojourn time distribution, in comparison with stopping at the data file 284 using an Erlang-2 sojourn time distribution.

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3*</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average cost</td>
<td>68.37</td>
<td>68.23</td>
<td>68.19</td>
<td>68.22</td>
<td>68.31</td>
<td>68.46</td>
<td>68.71</td>
<td>69.12</td>
<td>69.96</td>
</tr>
</tbody>
</table>

### 6.4 Mean Remaining Time in the Healthy State

Given parameter estimates \( r^* \), the mean remaining time (MRT) in the healthy state is fully determined by the posterior probability \( \Pi_n \). Using Eqs. (24-25), we determine the MRT given a gear tooth deterioration history up to the file 290 (RMS values) using an Erlang-2 sojourn time distribution. The MRT is predicted to be 4.52 hours at file 247. Using \( \Pi_n \) given in Figure 5, the MRT continuously decreases over time when the system is in the running period from data files 247 to 290, as shown in Figure 6. The computation of MRT using a HMM is considered as well, where the MRT is predicted to be 4.34 hours. It remains constant up to the file 284 when the system is in the unhealthy state. We note that the actual RT time from the data files 247 to 283 is 4.79 hours. Thus, the Erlang-2 model obtains better results than the HMM.

![Figure 5: Posterior Probabilities and the Control Limits for both Erlang-2 Distribution and Exponential Distribution](image)

![Figure 6: Mean Remaining Time in the Healthy State](image)

### 7. Conclusions

We have considered a Bayesian estimation and control problem for a partially observable deteriorating system subject to vibration monitoring. The system deterioration process has been described by a hidden, two state continuous-time semi-Markov process with an Erlang distribution in the healthy state. The objective was to determine the optimal maintenance policy that minimizes the long-run expected average cost per unit time. The state and observation processes have been modeled in the hidden semi-Markov framework and the maximum likelihood estimates of the model parameters have been obtained using the EM algorithm. We have shown that the multivariate Bayesian control chart can be used for both optimal maintenance decision-making as well as an effective cost-optimal early fault prediction scheme for the deteriorating system. The validation of the proposed methodology has been carried out using real vector vibration data. Also, the mean remaining life in the healthy state prediction formula has been developed. We have
compared the predicted MRT values in an Erlang-2 model, a HMM and the actual RT using real vibration data, and obtained a very good agreement. We have found that the results obtained using the hidden semi-Markov model have better performance than the results obtained using the hidden Markov model. The procedure is computationally efficient in terms of both parameter estimation and the SMDP optimization. An extension by considering CBM for a planetary gearbox system in a HSMM framework can be a suitable topic for future research.

Acknowledgement: The authors would like to thank the Natural Sciences and Engineering Research Council of Canada for the financial support under grant RGPIN 121384-11.

Appendix A

Theorem 1 Given $M$ observation histories, the pseudo log-likelihood function has the following decomposition:

$$
\sum_{i=1}^{M} \Omega_i(O,S|O_j,S_j) = \sum_{i=1}^{M} \Omega_i^{obs}(O|O_j,S_j) + \sum_{i=1}^{M} \Omega_i^{state}(S|O_j,S_j),
$$

(A. 1)

where

$$
\Omega_i^{obs}(O|O_j,S_j) = \langle b^i, \ln g \rangle,
$$

(A. 2)

$$
\Omega_i^{state}(S|O_j,S_j) = \langle a^i, c \rangle.
$$

(A. 3)

We denote vectors $a^i = (a_{0}^i, \ldots, a_{N}^i)$, $b^i = (b_{1}^i, \ldots, b_{N}^i)$, $c^i = (c_{1}^i, \ldots, c_{N}^i)$ and $g^i = (g_{0}^i, g_{1}^i, \ldots, g_{N}^i)$, where $g_{t}^i = g_{t}^i(y^i_t|\theta)$, with $n < t \leq (n+1)h$, $n = 0, \ldots, N_i - 1$. When $t > N_i h$, we write $g_{N_i}^i = g_{N_i}^i(y^i_{N_i}|\theta)$. Vectors $a^i$, $b^i$, $c^i$ and $g^i$ depend only on the fixed estimates $O_j, S_j$, which are given in the proof of Theorem 1 (see Appendix A). To simplify notation, for any vector $v = (v_0, v_1, \ldots, v_n)^t$, we denote $\ln v = (\ln v_0, \ln v_1, \ldots, \ln v_n)^t$. The inner product $(v, w) := v^t w$.

Proof. The pseudo log-likelihood can be written as follows:

$$
\sum_{i=1}^{M} \Omega_i(O,S|O_j,S_j) := \sum_{i=1}^{M} E_{(O_j,S_j)}(\ln L(O,S|y^i, t^i)|y^i).
$$

$$
= \sum_{i=1}^{M} \int_{0}^{\infty} \ln \left( g_{0}^i(y^i_t|\theta) f_{2}^i(t) \right) g_{0}^i(y^i_t|\theta) f_{3}^i(t) dt
$$

$$
= \sum_{i=1}^{M} \sum_{n=0}^{N_i-1} b_{N_i}^i \ln \left( g_{0}^i(y^i_n|\theta) \right) + b_{N_i}^i \ln \left( g_{0}^i(y^i_{N_i}|\theta) \right)
$$

(A. 4)

The first term in Eq. (A. 4) depends only on the observation parameter set $O$ and the second term in Eq. (A. 4) depends only on the state parameter set $S$. Thus, the observation term in Eq. (A. 1) can be written as follows:

$$
\Omega_i^{obs}(O|O_j,S_j) = \sum_{n=0}^{N_i-1} b_{N_i}^i \ln \left( g_{0}^i(y^i_{N_i}|\theta) \right)
$$

(A. 5)

where

$$
\begin{align*}
\hat{b}_{N_i} & = \frac{g_{0}^i(y^i_{N_i}|\theta)}{\int_{0}^{\infty} g_{0}^i(y^i_{N_i}|\theta) f_{3}^i(t) dt}, \\
\hat{b}_{N_i} & = \frac{g_{0}^i(y^i_{N_i}|\theta)}{\int_{0}^{\infty} g_{0}^i(y^i_{N_i}|\theta) f_{3}^i(t) dt}.
\end{align*}
$$

(A. 6)

and
\[
\int_0^\infty g_{ij}(\theta|f_j(t) dt = \sum_{n=0}^{N-1} g_{ij}(\theta|n)(R(j,n) - R(j,n+1)) + g_{ij}(\theta|N)R(j,N).
\] (A.7)

The reliability function \(R(j,n)\) represents the probability that the time to failure will be greater than a specific time \(nh\) considering parameter estimates \(\theta_j\). The state term in Eq. (A.4) can be written as follows:

\[
\dot{\Omega}^{\text{state}}(s|\Omega_j, \Sigma_j) = \sum_{n=0}^{N-1} a_{n} \cdot c_{n} + a_{N} \cdot c_{N},
\] (A.8)

where

\[
a_{n} = \frac{g_{ij}(\theta|n)}{\int_0^\infty g_{ij}(\theta|f_j(t) dt}, \quad n = 0, \ldots, N-1,
\]
\[
a_{N} = \frac{g_{ij}(\theta|N)}{\int_0^\infty g_{ij}(\theta|f_j(t) dt},
\]
\[
c_{n} = \int_{nh}^{(n+1)h} \ln(f_j(t)) f_j(t) dt, \quad n = 0, \ldots, N-1
\]
\[
c_{N} = \int_{Nh}^{\infty} \ln(f_j(t)) f_j(t) dt.
\] (A.10)

This completes the proof.

Appendix B

**Lemma 1** For \(M\) observation histories, the estimate of parameter \(r\) in the \((j+1)\)th step for given phase \(m\) is given by the following formula:

\[
r^{j+1} = m \sum_{n=0}^{N-1} e_1^j
\]
\[
a_1^j = (e_1^j, g_1^j),
\]
\[
y_1^j = (e_2^j, g_2^j),
\]
\[
e_1^j = \int_{nh}^{(n+1)h} t^m \cdot e^{-rt} dt, \quad n = 0, \ldots, N-1,
\]
\[
e_2^j = \int_{nh}^{(n+1)h} t^m \cdot e^{-rt} dt,
\]
\[
e_1^N = \int_{Nh}^{\infty} t^m \cdot e^{-rt} dt,
\]
\[
e_2^N = \int_{Nh}^{\infty} t^m \cdot e^{-rt} dt.
\]

where vectors \(e_1^j = [e_1^1, e_1^2, \ldots, e_1^N]\) and \(e_2^j = [e_2^1, e_2^2, \ldots, e_2^N]\) .

**Proof:**

\[
\Omega^{\text{state}}(s|\theta_j, \Sigma_j) = \sum_{n=0}^{N-1} a_{n} \int_{Nh}^{(n+1)h} (m \ln r + (m-1) \ln t - rt - \ln((m-1)!)) \frac{e_1^m e^{m-1} e^{-rt} dt}{(m_j-1)!} + a_{N} \int_{Nh}^{\infty} (m \ln r + (m-1) \ln t - rt - \ln((m-1)!)) \frac{e_1^m e^{m-1} e^{-rt} dt}{(m_j-1)!}.
\] (B.2)

For a given \(m\), the Lemma 1 can be obtained by taking partial derivative with respect to \(r\). The stationary point of state parameter \(r\) can be solved by \(\frac{\partial \Omega^{\text{state}}}{\partial r} = 0\). This completes the proof.

Appendix C

**Theorem 2** The observation parameter estimates \(\theta^{j+1} = (\mu_0^{j+1}, \gamma^{j+1}, \xi^{j+1}, \Sigma^{j+1})\) are given explicitly by:

\[
\mu_0^{j+1} = \frac{\sum_{i=1}^{M} h_i^{j+1} \cdot b_i^{j+1}}{\sum_{i=1}^{M} (b_i^{j+1})}, \quad \gamma^{j+1} = \frac{\sum_{i=1}^{M} h_i^{j+1} \cdot b_i^{j+1}}{\sum_{i=1}^{M} (b_i^{j+1})},
\]
\[
\xi^{j+1} = \frac{\sum_{i=1}^{M} x_i^{j+1} h_i^{j+1} \cdot b_i^{j+1}}{\sum_{i=1}^{M} x_i^{j+1} (b_i^{j+1})}, \quad \Sigma^{j+1} = \frac{\sum_{i=1}^{M} x_i^{j+1} h_i^{j+1} \cdot b_i^{j+1}}{\sum_{i=1}^{M} x_i^{j+1} (b_i^{j+1})},
\] (C.1)

where
A Comparison of Hidden Markov and Semi-Markov Modeling for a Deterioration System subject to Vibration Monitoring

\[ n_1 = (0, \sum_{i=1}^{\infty} y_{ni}, \cdots, \sum_{i=1}^{\infty} y_{ni}), \]

\[ n_2 = (\sum_{i=1}^{\infty} y_{ni}, \sum_{i=2}^{\infty} y_{ni}, \cdots, y_{N_n}, 0), \]

\[ n_3 = (0, \sum_{i=1}^{\infty} (y_{ni} - \mu_0)(y_{ni} - \mu_0)' \cdots \sum_{i=1}^{\infty} (y_{ni} - \mu_0)(y_{ni} - \mu_0)'), \]

\[ n_4 = (\sum_{i=1}^{\infty} (y_{ni} - \mu_1)(y_{ni} - \mu_1)' \cdots \sum_{i=1}^{\infty} (y_{ni} - \mu_1)(y_{ni} - \mu_1)'), \]

\[ d_1 = (0, 1, \cdots, N_n)', \quad d_2 = (N_n, \cdots, 1, 0)', \quad b_i = (b_{i1}, \cdots, b_{i_j})'. \quad (C.2) \]

References


Chen, Lin is a Ph.D. student in the Department of Mechanical and Industrial Engineering, University of Toronto. His research interests are in maintenance and reliability engineering, signal processing, and the development of fault detection and failure prevention schemes.

Makis, Viliam is a professor in the Department of Mechanical and Industrial Engineering, University of Toronto. His research and teaching interests are in the areas of quality assurance, stochastic OR modeling, maintenance, reliability, and production control with special interest in investigating the optimal operating policies for stochastic controlled systems. His recent contributions have been in the area of modeling and optimization of partially observable processes with applications in CBM and multivariate quality control. He has contributed also to the development of EMQ and other production models with inspections and random machine failures, joint SPC and APC for deteriorating production processes, scheduling of operations in FMS, reliability assessment of systems operating under varying conditions, and modeling and control of queuing systems. He was a founding member of the CBM Consortium at the University of Toronto in 1995. He is on the editorial board of several International Journals and has been the Editor of several international journals such as Advances in OR, ISRN Probability and Statistics and CEJE. He has also been on the Advisory Boards and on the Organizing and Technical Program Committees for several International Conferences. He is a Fellow of the International Society of Engineering Asset Management (FISEAM) and a senior member of IIE and ASQ.