Parameter and Residual Life Estimation for a Hidden Semi-Markov Model of a Deteriorating System

Akram Khaleghei
Dept. of Mechanical & Industrial Eng.
University of Toronto, 5 King's College Road,
Toronto, Ontario, Canada, M5S 3G8
akhalegh@mie.utoronto.ca

Viliam Makis
Dept. of Mechanical & Industrial Eng
University of Toronto, 5 King's College Road,
Toronto, Ontario, Canada, M5S 3G8
makis@mie.utoronto.ca

***Abstract*–This paper presents a new method for modeling and residual life estimation for a technical system subject to deterioration and random failure. The evolution of the system state process is modeled as a three-state hidden semi-Markov process where the sojourn times in both operational states follow a 2-phase Erlang distribution. The multivariate observation vectors related to the system state are available at discrete times through condition monitoring which provides partial information about the hidden condition of the system. Using the reference model approach and the residuals to represent the observation process, a new estimation procedure is developed and explicit formulas are derived for the system residual life estimation. A numerical example is presented to illustrate the whole procedure.**

***Keywords*–Expectation-maximization (EM) algorithm; Partially observable system; Phase-type distribution, Hidden semi-Markov model, Residual life estimation**

I. INTRODUCTION

Recently, due to the advances in data measurement and computer technology, it has become possible to implement effective condition monitoring (CM) systems for critical equipment which increase plant productivity.

The objective is to utilize the information obtained from CM for the assessment of the actual condition of the operating equipment without any unwanted disruption. The collected data carries only partial information about the unknown state of the equipment and the dimensionality of such data is typically very large, with lots of noise and cross and auto-correlation. Various approaches have been applied for processing and modeling of such data to develop prognostic procedures for systems subject to CM [2]. In this paper, we focus on the application of statistical modeling approaches. There are three main types of statistical models which have been widely used for prognostics using CM data based on indirectly observed state processes, namely stochastic filtering models [11], proportional hazard models (PHM) [7], and the hidden Markov models (HMM) [4]. Recently, some hidden semi-Markov models (HSMM) have also been developed for prognostics using CM data (see e.g. [8, 9]). Although HSMM have been successfully applied and provided promising results in many areas, the research on HSMM for fault prognosis based on CM data is very limited [10].

In this paper, we present a highly effective method for fault prognostics of systems subject to CM which is suitable for a wide range of deteriorating systems with random, observable failure. We model the deterioration process as a non-decreasing continuous time homogenous semi-Markov chain with two unobservable operational states and one observable failure state. The system is considered to be in the healthy state while degradation is below a critical level. The maintenance actions are only initiated when the system is in the warning state which indicates that the system experiences severe degradation which can cause failure. The two state HMM has been proposed in the recent studies which used real data obtained from CM such as spectrometric oil data [7] and vibration data [12] for condition based maintenance (CBM). The authors showed that considering the HMM with 2 states ("in-control" and "out-of-control") is sufficient for an early fault detection and condition based maintenance. It has also been noted in those papers that some real data histories were very short, indicating the occurrence of failure in the healthy state which led to the development of a HMM with the possibility of a direct transition to failure state from a healthy state. However, HMMs generally assume that the sojourn time in each state is exponentially distributed which is not always a realistic assumption. In this paper, we assume that the sojourn times in both operational states have 2-phase Erlang distributions. While the system is operational, vector data that is stochastically related to the hidden system state is obtained through CM at regular sampling epochs.We assume that two types of data histories are available: histories that end with observable system failure, and censored data histories that end when the system has been suspended from operation but has not failed. We apply the Expectation-Maximization (EM) algorithm to estimate the model parameters. Although some researchers have investigated parameter estimation for partially observable systems in both HMM (see e.g. [5, 14]) and the HSMM framework (see e.g. [1, 3]), to our knowledge, there are limited references that considered both non-exponential sojourn times for continuous-time deterioration processes and also the failure information.

In fact, [3] is the only reference where an estimation procedure was developed by considering a simple Erlang(2,λ) distribution for the sojourn time in state 1 and an exponential distribution in state 2, plus the failure information. Our paper is different from [3] in several important aspects. First, we consider 2-phase Erlang sojourn time distribution for both operational states which is more appropriate for modeling real systems. Second, we introduce the transition probability as an unknown state parameter and derive a close form re-estimation formula for it, whereas in [3] the parameter was assumed to be a function of other model parameters and it was not estimated. We derive the explicit formulas for the conditional reliability function (RF) and the MRL function in terms of the posterior probability statistic which has already been shown to be a sufficient statistic for decision-making [6]. We have found that both MRL and RF indicators immediately identify the time when the system starts experiencing severe degradation. For the performance evaluation and illustration of the ability of the developed model for fault prediction, we compare the estimated conditional MRL and RF with those estimated using HMM. Furthermore, to test the accuracy of the proposed model to detect the change for different value of shift size in the mean vector of observations, the small, medium and large size of shifts have been investigated. The relative percentage of deviation is used to compare the capability of the proposed HSMM with the HMM for fault prognostic. It is found that the proposed HSMM demonstrates different performances for different shift sizes and on average it produces 55.66% lower RPDs in all the three different scenarios when compared with HMM. Therefore, we can confidently assert that HSMM provides considerably more precise MRL estimates in all the three cases investigated. We have also observed that, failure to accurately estimate the MRL is propagated with the new data when the conditional RF of the system is calculated using HMM.

II. HSM MODELING AND PARAMETER ESTIMATION

We assume that a system's operating condition can be categorized into two operational states known as healthy state (state 1) and unhealthy state (state 2), and one observable failure state (state 3), where the sojourn time in operating state i follows a 2-phase Erlang distribution with parameters $(2, \lambda_i)$, for $i = 1, 2$. We model the state process $(X_t : t \in \mathbb{R}^+)$ as a continuous time homogeneous semi-Markov chain with state space $\mathcal{X} = \{1, 2, 3\}$. The machine starts working in a healthy state and it can make transitions from healthy state to unhealthy state with the probability p_{12} due to degradation, or from healthy state to failure state with probability p_{13}, where $p_{12} + p_{13} = 1$. The system condition is monitored at equidistant sampling epochs $\Delta, 2\Delta, \ldots$ for $\Delta \in (0, +\infty)$ and vector data $Y_1, Y_2, \ldots \in \mathbb{R}^d$ is collected at these times which represent partial information about the system state. It is assumed that the observations have d-dimensional normal distribution $N_d(\mu_i, \Sigma_i)$, and are conditionally independent given the system state i.e. $Y_n | X_{n\Delta} = i \sim f_i(y_n) \overset{iid}{\sim} N_d(\mu_i, \Sigma_i)$ for $i = 1, 2$ where $\mu_i \in \mathbb{R}^d$ and $\Sigma_i \in \mathbb{R}^d \times \mathbb{R}^d$ are the unknown observation process parameters. These assumptions are reasonable in real application once appropriate data pre-processing methods have been applied (see e.g.[13]). We also assume that two types of data histories are available: failure histories and suspension histories. Using the phase-type property of Erlang distribution, we enlarge the state space. Let the new state space be $\mathcal{Z} = \{1, 2, 3, 4, 5\}$, where states $\{1, 2\}$ denote that the machine is working in a healthy condition and states $\{3, 4\}$ indicate that machine is operating in a unhealthy condition. State $\{5\}$ represents the observable failure (absorbing) state. We then model the state process $(Z_t : t \in \mathbb{R}^+)$ as a continuous time homogeneous Markov chain with state space $\mathcal{Z}$. The instantaneous transition rates $Q = (q_{ij})_{i,j \in \mathcal{Z}}$ for the state process from state i to state j are given by:

$$Q = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & 0 & 0 \\ 0 & -\lambda_1 & p_{12}\lambda_1 & 0 & p_{13}\lambda_1 \\ 0 & 0 & -\lambda_2 & \lambda_2 & 0 \\ 0 & 0 & 0 & -\lambda_2 & \lambda_2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and otherwise $q_{ij} = 0$. The transition probability matrix $\mathbb{P}(t) = \left(P_{ij}(t)\right)_{i,j \in \mathcal{Z}}$ for state process Z_t can be obtained by solving the Kolmogorov backward differential equations giving the following results:

$$P_{11}(t) = e^{-\lambda_1 t}, \qquad P_{12}(t) = \lambda_1 t e^{-\lambda_1 t}$$
$$P_{13}(t) = \Big(\frac{e^{-\lambda_2 t} - e^{-\lambda_1 t}}{(\lambda_1-\lambda_2)^2} + t\frac{e^{-\lambda_1 t}}{\lambda_2-\lambda_1}\Big)\lambda_1^2 p_{12}$$
$$P_{14}(t) = \Big(2\frac{e^{-\lambda_1 t} - e^{-\lambda_2 t}}{(\lambda_1-\lambda_2)^3} + t\frac{e^{-\lambda_1 t} + e^{-\lambda_2 t}}{(\lambda_1-\lambda_2)^2}\Big)\lambda_1^2\lambda_2 p_{12}$$
$$P_{24}(t) = \Big(\frac{e^{-\lambda_1 t} - e^{-\lambda_2 t}}{(\lambda_1-\lambda_2)^2} + t\frac{e^{-\lambda_2 t}}{\lambda_1-\lambda_2}\Big)\lambda_1\lambda_2 p_{12},$$
$$P_{33}(t) = e^{-\lambda_2 t}, P_{34}(t) = \lambda_2 t e^{-\lambda_2 t}, P_{44}(t) = e^{-\lambda_2 t}$$
$$P_{i5}(t) = 1 - \sum_{j=i}^{5} P_{ij}(t) \text{ for } i = 1,\dots,5$$

Let $\xi = \inf\{t > 0 | X_t = 3\}$ denote the time to failure and τ_1 represents the sojourn time in healthy state. Further suppose that we have collected M failure and N suspension histories. Failure history F_i is assumed to be of the form $\vec{Y}_i = (y_1^i, \dots, y_{T_i}^i)$ for $T_i\Delta < \xi_i \le (T_i+1)\Delta$, and suspension history S_j is assumed to be of the form $\vec{Y}_j = (y_1^j, \dots, y_{T_j}^j)$ for $\xi > T_j\Delta$. Let $C = \{F_1, \dots, F_M, S_1, \dots, S_N\}$ represent all the observable data set and $\bar{C} = \{\bar{F}_1, \dots, \bar{F}_M, \bar{S}_1, \dots, \bar{S}_N\}$ denotes the whole complete data set where each failure history and suspension history has been compounded with the unobservable sample path information of the hidden state process. Let $\Lambda = (\lambda_1, \lambda_2, p_{12})$ and $\Psi = (\mu_1, \Sigma_1, \mu_2, \Sigma_2)$ be the sets of unknown state and observation parameters. For the complete data histories, the associated likelihood function is given by:

$$L(\Lambda, \Psi|\bar{C}) = \prod_{i=1}^{M} L_{\bar{F}_i(\Lambda,\Psi)} \prod_{j=1}^{N} L_{\bar{S}_j(\Lambda,\Psi)} \tag{1}$$

where,

$$L_{\bar{F}}(\Lambda,\Psi) = \begin{cases} g_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,w) f_{\tau_1|\xi}(w|t) f_\xi(t) & w < t \\ g_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,t) m_{\tau_1|\xi}(t|t) f_\xi(t) & w = t \end{cases} \tag{2}$$

represents a complete likelihood function for a single failure history where $m_{\tau_1|\xi}(t|t) = P(\tau_1 = t|\xi = t)$ is a conditional probability function of τ_1 given ξ. Also $L_{\bar{S}}(\Lambda,\Psi)$ in Eq. (1) represents a complete likelihood function of a suspension history given by:

$$L_{\bar{S}}(\Lambda,\Psi) = \; = \; g_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,w) h_{\xi|\tau_1}(t|w) f_{\tau_1}(w) \tag{3}$$

where $h_{\xi|\tau_1}(t|w)$ is the conditional reliability of ξ given τ_1. The $g_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,w)$ in Eqs. (2) and (3) denotes the conditional density of the observation process given failure time and sojourn time. For any $w \in ((k-1)\Delta, k\Delta], k = 1, 2, \dots, T$:

$$g_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,w) = g_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,k\Delta)$$
$$= \frac{exp\Big(-\frac{1}{2}\sum_{n=1}^{k-1}(y-\mu_1)'\Sigma_1^{-1}(y-\mu_1) - \frac{1}{2}\sum_{n=k}^{T}(y-\mu_2)'\Sigma_2^{-1}(y-\mu_2)\Big)}{\sqrt{(2\pi)^{Td}|\Sigma_1|^{k-1}|\Sigma_2|^{T-k+1}}} \tag{4}$$

and for any $w > T\Delta$:

$$g_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,w) = g_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,t) \tag{5}$$
$$= \frac{1}{\sqrt{(2\pi)^{Td}|\Sigma_1|^T}} exp\Big(-\frac{1}{2}\sum_{n=1}^{T}(y-\mu_1)'\Sigma_1^{-1}(y-\mu_1)\Big)$$

In the next theorems, we derive explicit formulas for the density of ξ and also for the joint density of (ξ, τ_1) which will be used later to develop the pseudo likelihood function.

Theorem 1. *Let $F_1(t)$ and $F_2(t)$ be the cumulative distribution functions of the sojourn times in healthy and unhealthy states, respectively. For each $t \in \mathbb{R}^+$, the density function of time to failure is given by:*

$$f_\xi(t) = p_{12}\mathcal{L}_s^{-1}\Big(f_1^\star(s).f_2^\star(s)\Big) + p_{13}f_1(t)$$

where $f^\star(s) = \int_{t=0}^{\infty} e^{-st}f(t)dt$ denotes the Laplace transform (LT) of $f(t)$ and $\mathcal{L}_s^{-1}(f(s))$ is the inverse of LT of $f(s)$.

Proof. Let's assume that S_1 represents the system state at time τ_1.

$$\begin{aligned} P(\xi \le t) &= p_{12}P(\xi \le t|S_1 = 2) + p_{13}P(\xi \le t|S_1 = 3) \\ &= p_{12}\int_{u=0}^{t} P(\xi \le t|\tau_1 = u, S_1 = 2)dF_1(u) \\ &+ p_{13}\int_{u=0}^{t} P(\xi \le t|\tau_1 = u, S_1 = 3)dF_1(u) \\ &= p_{12}\int_{u=0}^{t} P(X_{t-u} = 3|X_1 = 2)dF_1(u) \\ &+ p_{13}\int_{u=0}^{t} dF_1(u) \\ &= p_{12}\int_{u=0}^{t} F_2(t-u)f_1(u)du + p_{13}F_1(t) \end{aligned} \tag{6}$$

This expression cannot be integrated analytically for most of lifetime distributions such as Weibull, Lognormal and Erlang distributions. This equation may be integrated by taking the Laplace transform of both sides. By taking the Laplace transform of both sides of Eq. (6), and changing the order of integration in the first term involving the double integrals, the following is obtained:

$$\int_{t=0}^{\infty} e^{-st} P(\xi \le t)dt = p_{12}\Big(\int_{t=0}^{\infty} e^{-st}\int_{u=0}^{t} F_2(t-u) f_1(u)du\Big)dt + p_{13}\int_{t=0}^{\infty} e^{-st}F_1(t)dt$$

$$F_\xi^\star(s) = p_{12}\int_{u=0}^{\infty}\int_{t=u}^{\infty} e^{-st}F_2(t-u)f_1(u)dtdu + p_{13}F_1^\star(s)$$

Performing a change of variable by defining $y = t - u$, the following is obtained:

$$\begin{aligned} F_\xi^\star(s) &= p_{12}\int_{u=0}^{\infty}\int_{y=0}^{\infty} e^{-s(y+u)}F_2(y)f_1(u)dydu \\ &\quad + p_{13}F_1^\star(s) \\ &= p_{12}\int_{u=0}^{\infty} e^{-su}f_1(u)du\int_{y=0}^{\infty} e^{-sy}F_2(y)dy \\ &\quad + p_{13}F_1^\star(s) = p_{12}f_1^\star(s)F_2^\star(s) + p_{13}F_1^\star(s) \quad (7) \end{aligned}$$

Using the property of Laplace transform of cumulative distribution function i.e. $F^\star(s) = \frac{f^\star(s)}{s}$, Eq. (7) can be written as,

$$f_\xi^\star(s) = p_{12}f_1^\star(s).f_2^\star(s) + p_{13}f_1^\star(s)$$

Inverting Laplace transform of the above equation gives the desired result,

$$f_\xi(t) = p_{12}\mathcal{L}_s^{-1}\Big(f_1^\star(s).f_2^\star(s)\Big) + p_{13}f_1(t)$$

□

Theorem 2. *For all non-negative* $0 < w < t$*, the joint density of* ξ *and* τ_1 *is given by,*

$$\begin{aligned} f_{(\xi,\tau_1)}(t,w) &= \mathcal{L}_{(s,v)}^{-1}\Big(p_{12}f_1^\star(s+v)f_2^\star(s) \\ &\quad + p_{13}\frac{v}{s+v}f_1^\star(v+s)\Big) \end{aligned}$$

Proof.

$$\begin{aligned} P(\xi \le t, \tau_1 \le w) &= p_{12}P(\xi \le t, \tau_1 \le w|S_1 = 2) \\ &\quad + p_{13}P(\tau_1 \le w, \xi \le t|S_1 = 3) \\ &= p_{12}\int_{u=0}^{w} P(\xi \le t|\tau_1 = u, S_1 = 2)dF_1(u) \\ &\quad + p_{13}\int_{u=0}^{w} P(\xi \le t|\tau_1 = u, S_1 = 3)dF_1(u) \\ &= p_{12}\int_{u=0}^{w} P(X_{t-u} = 3|X_1 = 2)f_1(u) \\ &\quad + p_{13}\int_{u=0}^{w} f_1(u) \end{aligned}$$

$$= p_{12}\int_{u=0}^{w} F_2(t-u)f_1(u)du + p_{13}F_1(w) \quad (8)$$

For the joint distribution function $f(t,w)$, the Laplace transform is defined as:

$$\mathcal{L}(f(t,w)) = f^\star(s,v) = \int_{t=0}^{\infty}\int_{w=0}^{\infty} e^{-st-vw}f(t,w)dwdt.$$

Taking the Laplace transform of both sides of Eq. (8), we have,

$$\begin{aligned} F^\star(s,v) &= p_{12}\int_{t=0}^{\infty}\int_{w=0}^{t}\int_{u=0}^{w} e^{-st-vw}F_2(t-u)f_1(u)dudwdt \\ &\quad + p_{13}\int_{t=0}^{\infty}\int_{w=0}^{t} e^{-st-vw}F_1(w)dwdt \quad (9) \end{aligned}$$

Changing the order of integration in the first and second terms and performing a change of variable by defining $y = t - u$, the following is obtained:

$$\begin{aligned} F^\star(s,v) &= -p_{12}\Big(\int_{u=0}^{\infty}\int_{w=0}^{u}\int_{y=0}^{\infty} e^{-s(y+u)-vw}F_2(y) \\ &\quad f_1(u)dydwdu\Big) + p_{13}\int_{w=0}^{\infty} e^{-vw}F_1(w)\int_{t=w}^{\infty} e^{-st}dtdw \\ &= -p_{12}\int_0^{\infty}\Big(\frac{1}{v}(1-e^{-uv})\Big)e^{-su}f_1(u)du\int_0^{\infty} e^{-sy}F_2(y)dy \\ &\quad + \frac{p_{13}}{s}\int_{w=0}^{\infty} e^{-w(v+s)}F_1(w)dw \\ &= -\frac{p_{12}}{v}\big(f_1^\star(s) - f_1^\star(s+v)\big)F_2^\star(s) + \frac{p_{13}}{s}F_1^\star(v+s) \end{aligned}$$

Using the property of Laplace transform of cumulative joint distribution function i.e. $F^\star(s,v) = \frac{1}{sv}f^\star(s,v)$, above equation can be written as,

$$f_{(\xi,\tau_1)}^\star(s,v) = -p_{12}\big(f_1^\star(s) - f_1^\star(s+v)\big)f_2^\star(s) + p_{13}\frac{vf_1^\star(v+s)}{s+v}$$

Finally, taking inverse Laplace transform of above equation we have:

$$f_{(\xi,\tau_1)}(s,v) = p_{12}\mathcal{L}_{(s,v)}^{-1}\left(f_1^\star(s+v)f_2^\star(s) + p_{13}\frac{v}{s+v}f_1^\star(v+s)\right)$$

□

Using the results of the Theorem 1 and 2, the distributional property of the time to failure of the proposed model is given by:

$$\begin{aligned} f_\xi(t) &= p_{12}\lambda_1^2\lambda_2^2\Big(\frac{2(e^{-\lambda_1 t} - e^{-\lambda_2 t})}{(\lambda_1-\lambda_2)^3} + \frac{t(e^{-\lambda_1 t} + e^{-\lambda_2 t})}{(\lambda_1-\lambda_2)^2}\Big) \\ &\quad + p_{13}\lambda_1^2 te^{-\lambda_1 t} \quad (10) \end{aligned}$$

and for all $0 < w < t$, the joint density of τ_1 and ξ is given by:

$$f_{(\xi,\tau_1)}(t,w) = p_{12}\lambda_1^2\lambda_2^2 e^{-\lambda_2(t-w)-\lambda_1 w}(t-w)w \quad (11)$$

Details can be found in Appendix A. Furthermore, the conditional reliability function of ξ given $\tau_1 = w$ for $w \le t$ is given by,

$$h_{\xi|\tau_1}(t|w) = p_{12}e^{-\lambda_2(t-w)}(1+\lambda_2(t-w)) \quad (12)$$

and $h_{\xi|\tau_1}(t|w) = 1$ for all $w > t$. Next, we evaluate the E-step of the EM algorithm by performing the expectation of the log of the likelihood function given in Eq. (1):

$$Q(\Lambda,\Psi|\hat{\Lambda},\hat{\Psi}) = Q^{state}(\Lambda|\hat{\Lambda},\hat{\Psi}) + Q^{obs}(\Psi|\hat{\Lambda},\hat{\Psi}) \quad (13)$$

Since the pseudo log-likelihood function can be decomposed as a function of state parameter and residual observation parameters (details can be found in Appendix B, the M-step can be carried out separately for the state and residual observation parameters. We solve for the stationary point of the observation parameter by setting:

$$\frac{\partial Q^{obs}(\Psi|\hat{\Lambda},\hat{\Psi})}{\partial \mu_1} = \frac{\partial Q^{obs}(\Psi|\hat{\Lambda},\hat{\Psi})}{\partial \mu_2} = \mathbf{0}$$
$$\frac{\partial Q^{obs}(\Psi|\hat{\Lambda},\hat{\Psi})}{\partial \Sigma_1^{-1}} = \frac{\partial Q^{obs}(\Psi|\hat{\Lambda},\hat{\Psi})}{\partial \Sigma_2^{-1}} = \mathbf{0}$$

After some algebra, it is not difficult to check that there is a unique stationary point $\hat{\Psi} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\Sigma}_1, \hat{\Sigma}_2)$ of the pseudo log-likelihood function given by:

$$\hat{\mu}_1 = \frac{\sum_{i=1}^M \langle \hat{\mathbf{C}}^i, \mathbf{n}_1^i\rangle + \sum_{j=1}^N \langle \hat{\mathbf{D}}^j, \mathbf{n}_1^j\rangle}{\sum_{i=1}^M \langle \hat{\mathbf{C}}^i, \mathbf{s}_1^i\rangle + \sum_{j=1}^N \langle \hat{\mathbf{D}}^j, \mathbf{s}_1^j\rangle}$$
$$\hat{\Sigma}_1 = \frac{\sum_{i=1}^M \langle \hat{\mathbf{C}}^i, \mathbf{n}_3^i\rangle + \sum_{j=1}^N \langle \hat{\mathbf{D}}^j, \mathbf{n}_3^j\rangle}{\sum_{i=1}^M \langle \hat{\mathbf{C}}^i, \mathbf{s}_1^i\rangle + \sum_{j=1}^N \langle \hat{\mathbf{D}}^j, \mathbf{s}_1^j\rangle}$$
$$\hat{\mu}_2 = \frac{\sum_{i=1}^M \langle \hat{\mathbf{C}}^i, \mathbf{n}_2^i\rangle + \sum_{j=1}^N \langle \hat{\mathbf{D}}^j, \mathbf{n}_2^j\rangle}{\sum_{i=1}^M \langle \hat{\mathbf{C}}^i, \mathbf{s}_2^i\rangle + \sum_{j=1}^N \langle \hat{\mathbf{D}}^j, \mathbf{s}_2^j\rangle}$$
$$\hat{\Sigma}_2 = \frac{\sum_{i=1}^M \langle \hat{\mathbf{C}}^i, \mathbf{n}_4^i\rangle + \sum_{j=1}^N \langle \hat{\mathbf{D}}^j, \mathbf{n}_4^j\rangle}{\sum_{i=1}^M \langle \hat{\mathbf{C}}^i, \mathbf{s}_2^i\rangle + \sum_{j=1}^N \langle \hat{\mathbf{D}}^j, \mathbf{s}_2^j\rangle}$$

where $\mathbf{n}_1^i = (0, y_1^i, \sum_{n=1}^2 y_n^i, \ldots, \sum_{n=1}^{T_i} y_n^i)'$, $\mathbf{n}_2^i = (\sum_{n=1}^{T_i} y_n^i, \sum_{n=2}^{T_i} y_n^i, \ldots, y_{T_i}^i, 0)'$, $\mathbf{s}_1^i = (0, 1, \ldots, T_i)$, and $\mathbf{s}_2^i = (T_i, \ldots, 1, 0)'$. Using Lagrangian multiplier, the following formulas can be obtained:

$$\hat{p}_{12} = \frac{\sum_{i=1}^M \hat{\alpha}_1^i + \sum_{j=1}^N \hat{\beta}_1^j}{\sum_{i=1}^M (\hat{\alpha}_1^i + \hat{\eta}^i) + \sum_{j=1}^N \hat{\beta}_1^j},$$

$$\hat{p}_{13} = \frac{\sum_{i=1}^M \hat{\eta}^i}{\sum_{i=1}^M (\hat{\alpha}_1^i + \hat{\eta}^i) + \sum_{j=1}^N \hat{\beta}_1^j}$$

Moreover, λ_1 and λ_2 can be obtained by using:

$$\hat{\lambda}_1 = \frac{2\sum_{i=1}^M (\hat{\alpha}_1^i + \hat{\eta}^i) + 2\sum_{j=1}^N (\hat{\beta}_1^j + \hat{\nu}^j)}{\sum_{i=1}^M (\hat{\alpha}_3^i + \hat{\eta}^i t) + \sum_{j=1}^N (\hat{\beta}_3^j + \hat{q}^j)},$$
$$\hat{\lambda}_2 = \max_{\lambda_2} \sum_{i=1}^M (2\hat{\alpha}_1^i \ln \lambda_2 - \hat{\alpha}_2^i \lambda_2) + \sum_{j=1}^N \Big(\frac{\langle \hat{\mathbf{g}}, \mathbf{a}_5\rangle^j}{\hat{f}^j} - \hat{\beta}_2^j \lambda_2 \Big)$$

where $\hat{\alpha}_i = \frac{\langle \hat{\mathbf{g}}, \hat{\mathbf{d}}_i\rangle}{\hat{d}}$ and $\hat{\beta}_i = \frac{\langle \hat{\mathbf{g}}, \hat{\mathbf{a}}_i\rangle}{\hat{f}}$ for $i = 1, 2, 3$.

III. MEAN RESIDUAL LIFE ESTIMATION

Let $\Pi_n(i) = P(Z_{n\Delta} = i | y_1, \ldots, y_n, \xi > n\Delta)$ denote the posterior probability that the system is in state i at the nth decision epoch for $i \in \mathcal{Z}$ where $\Pi_0(1) = 1$. Using Bayes' rule for $n \ge 1$ we have,

$$\Pi_n(i) = \frac{g(y_n | Z_{n\Delta} = i, \xi > n\Delta, y_{n-1}, \Pi_{n-1}) C_{n-1}^i}{f_1(y_n)\sum_{j=1}^2 C_{n-1}^j + f_2(y_n)\sum_{j=3}^4 C_{n-1}^j}$$

where,

$$C_{n-1}^j = \sum_{i=1}^j P_{i,j}(\Delta)\Pi_{n-1}(i) \quad j = 1, 2$$

$$C_{n-1}^j = \sum_{i=1}^j P_{i,j}(\Delta)\Pi_{n-1}(i) \quad j = 3, 4$$

$$g(y_n | Z_{n\Delta} = i, \xi > n\Delta, y_{n-1}) = \begin{cases} f_1(y_n) & \text{if } i \in \{1, 2\} \\ f_2(y_n) & \text{if } i \in \{3, 4\} \end{cases}$$

Let $\vec{\Pi}_n = (\Pi_n(1), \ldots, \Pi_n(4))$ denote the vector of posterior probability at the nth decision epoch. For any $t \ge 0$, the conditional RF and the MRL at the nth decision epoch are given by,

$$\begin{aligned} R(t|\vec{\Pi}_n) &= P(\xi > n\Delta + t | \xi > n\Delta, Y_n = y_n, \vec{\Pi}_n) \\ &= \sum_{1\le i\le 4}\sum_{i\le j\le 4} P(Z_{n\Delta+t} = j | Z_{n\Delta} = i, \xi > n\Delta, \\ &\quad y_n, \vec{\Pi}_{n-1}) \times P(Z_{n\Delta} = i | \xi > n\Delta, y_n, \vec{\Pi}_{n-1}) \\ &= \sum_{i=1}^4 (1 - P_{i5}(t))\Pi_n(i) \\ \mu_{n\Delta} &= E(\xi - n\Delta | \xi > n\Delta, Y_n = y_n, \vec{\Pi}_n) \\ &= \int_0^\infty R(t|\vec{\Pi}_n)dt = (\frac{2}{\lambda_1} + \frac{2}{\lambda_2}p_{12})\Pi_n(1) \\ &+ (\frac{1}{\lambda_1} + \frac{2}{\lambda_2}p_{12})\Pi_n(2) + \frac{2}{\lambda_2}\Pi_n(3) + \frac{1}{\lambda_2}\Pi_n(4) \end{aligned}$$

Table 1: *Optimal parameter estimates using the EM algorithm*

Parameters	Initial values	1st iteration	2nd iteration	Final estimation
λ_1	.15	.20	.23	.25
λ_2	.50	.66	.69	.67
p_{12}	.65	.78	.82	.89
p_{13}	.35	.12	.18	.11
μ_1	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} .51 \\ .34 \end{pmatrix}$	$\begin{pmatrix} .46 \\ .28 \end{pmatrix}$	$\begin{pmatrix} .29 \\ -.10 \end{pmatrix}$
μ_2	$\begin{pmatrix} 1.5 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1.91 \\ 1.18 \end{pmatrix}$	$\begin{pmatrix} 2.05 \\ 1.25 \end{pmatrix}$	$\begin{pmatrix} 1.91 \\ 1.47 \end{pmatrix}$
Σ_1	$\begin{pmatrix} .50 & .25 \\ .25 & 1.00 \end{pmatrix}$	$\begin{pmatrix} .58 & .33 \\ .33 & 1.48 \end{pmatrix}$	$\begin{pmatrix} .67 & .38 \\ .38 & 1.56 \end{pmatrix}$	$\begin{pmatrix} 1.11 & .48 \\ .48 & 1.75 \end{pmatrix}$
Σ_2	$\begin{pmatrix} 1 & .5 \\ .5 & 1.5 \end{pmatrix}$	$\begin{pmatrix} 1.31 & .87 \\ .87 & 1.77 \end{pmatrix}$	$\begin{pmatrix} 1.64 & 1.07 \\ 1.07 & 1.92 \end{pmatrix}$	$\begin{pmatrix} 2.08 & 1.47 \\ 1.47 & 2.37 \end{pmatrix}$
$Q \times 10^3$	-	-1.45	-1.35	-1.21
Time (sec)	-	4.40	7.47	15.53

IV. Numerical example

A numerical example to illustrate the entire estimation procedure is presented in this section. Let's assume that the system deterioration follows a continuous-time homogenous semi-Markov chain $(X_t : t \in \mathbb{R}^+)$ with state space $\mathcal{X} = \{1, 2, 3\}$. The states 1 and 2 are unobservable, representing the healthy and unhealthy operational states respectively. The sojourn time in healthy and unhealthy state follows a 2 phase Erlang distribution with parameter λ_1 andλ_2, respectively. The system can make transitions from healthy state to warning state with the probability p_{12} or from healthy state to failure state with the provability p_{13}, where $p_{12} + p_{13} = 1$. The state parameters and the transitions probabilities are given by,

$$\lambda_1 = .3, \quad \lambda_2 = .8, \quad p_{12} = .80, \quad p_{13} = .20$$

At equidistant sampling times $n\Delta, \Delta = .1$, the observations $\{Y_n, n = 1, 2, \ldots\}$ are collected through condition monitoring and Y_n is assumed to follow 2-dimensional normal distribution $N_2(\mu_1, \Sigma_1)$ or $N_2(\mu_2, \Sigma_2)$, depending on whether the system is in the healthy or unhealthy state, where,

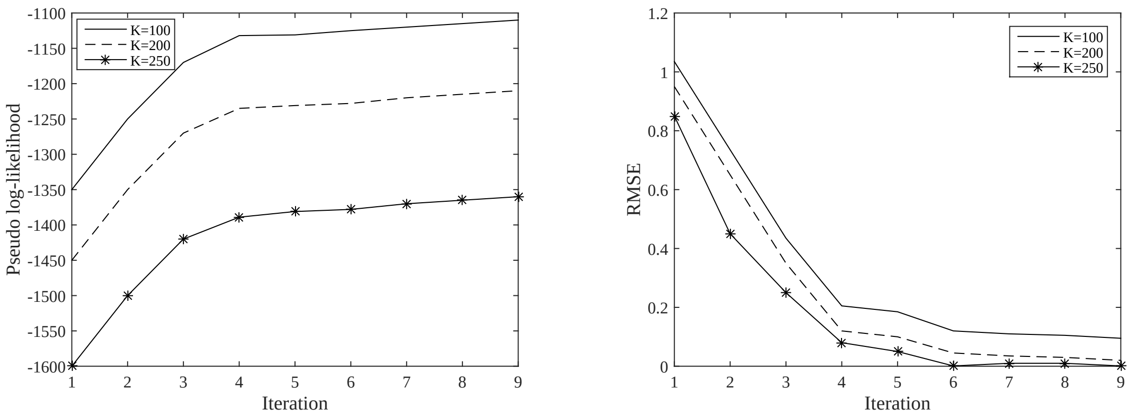
$$\mu_1 = \begin{pmatrix} .2 \\ -.1 \end{pmatrix}, \quad \Sigma_1 = \begin{pmatrix} 1.3 & .5 \\ .5 & 1.8 \end{pmatrix},$$
$$\mu_2 = \begin{pmatrix} 2.0 \\ 1.5 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 2 & 1.5 \\ 1.5 & 2.5 \end{pmatrix},$$

Using these parameters, $K = 200$ number of failure and suspension are generated. Running the proposed estimation procedure on the generated data with a stopping criterion of $\| (\Lambda_{n+1}, \Psi_{n+1}) - (\Lambda_n, \Psi_n) \| \leq 10^{-6}$, the parameters are estimated. Table 1 shows the estimated parameters in each iteration. The EM algorithm takes on average 11.53 second, which is extremely fast for off-line computations. This is an attractive feature for real applications. Before using the parameters' estimates for reliability prediction, we evaluate the performance of the estimation procedure by its ability to increase the pseudo log-likelihood function. As shown in Figure 1.a, the pseudo log-likelihood function is increasing in each iteration and also as the number of observation increases the corresponding pseudo log-likelihood's value decreases in all the investigated scenarios. We next evaluate the correctness of the parameter estimates for different number of collected histories using root mean square error (RMSE) measurement. For n number of unknown parameters, the RMSE measures the difference between the estimated values $\hat{\theta}$ and the actual values θ and can be obtained by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{\theta}_i - \theta_i)^2}{n}}$$

As shown in Figure 1.2 , the RMSE values is relatively higher for $K = 100$. By increasing the number of observation histories, the estimated error decreases.

Now that the parameters' estimates have been obtained, the model can be used for wear prediction based on conditional MRL and RF. For this purpose, a new group of 100 failure histories using the true model parameters is simulated. The conditional mean residual life and reliability functions for each failure history are then computed at each sampling epoch using the optimal parameter estimates. Estimated reliability at each sampling epoch for the particular failure history is provided in Figure 2.a. For the performance evaluation, we also compare the estimated conditional MRL and RF with those of estimated from hidden Markov modeling. The conditional MRL and RF for HMM can be estimated using:



a.Performance of the estimation procedure

b. The effect of number of histories on RMSE

Figure 1: *Performance evaluation of the parameter estimates*

$$\mu_{n\Delta} = \frac{\lambda_2 + p_{12}\lambda_1 + \Pi_n(p_{13}\lambda_1 - \lambda_2)}{\lambda_1\lambda_2}$$

$$R(t|\Pi_n) = \frac{\Big(\lambda_1 e^{-\lambda_1 t} - \lambda_2 e^{-\lambda_2 t} + (\lambda_2 + p_{12}\lambda_1 + \Pi_n(p_{13}\lambda_1 - \lambda_2))(e^{-\lambda_2 t} - e^{-\lambda_1 t})\Big)}{\lambda_1 - \lambda_2}$$

where Π_n is the probability of being in the unhealthy state at n^{th} decision epoch. As Figure 2.a shows, one epoch after the change occurred the computed reliability values dropped significantly for both models. This suggests that there is a high probability that the system working condition has changed and therefore an immediate preventive maintenance should be considered to avoid producing non-conforming products and possible production shutdowns due to system failure. The estimated MRL for both models at each sampling epoch for the investigated failure history are plotted in Figure 2.b. The MRL statistics for both models drop noticeably after the change occurred with one sampling interval delay. Both trends suggest that the working state has experienced a status change from healthy to unhealthy condition around time 18.3 and full system inspection and preventive maintenance might be required. The small drops in both estimated RF or MRL before the occurrence of the change tend to recover quickly in next sampling epochs.

To test the accuracy of the proposed model to detect the change for different values of μ_2, three different values for μ_2 are investigated and the MRL statistics at the last sampling epoch before failure occurrences are computed for all the simulated failure histories using the above mentioned procedure. Relative percentage deviation (RPD) as a common performance measure has been used to compare the capability of the proposed HSMM model with HMM for fault prognostic using Bayesian MRL statistics. RPD is obtained by the following formula,

$$RPD = \frac{1}{n}\sum \left|\frac{ARL - MRL_{estimate}}{ARL}\right| \times 100, \qquad (14)$$

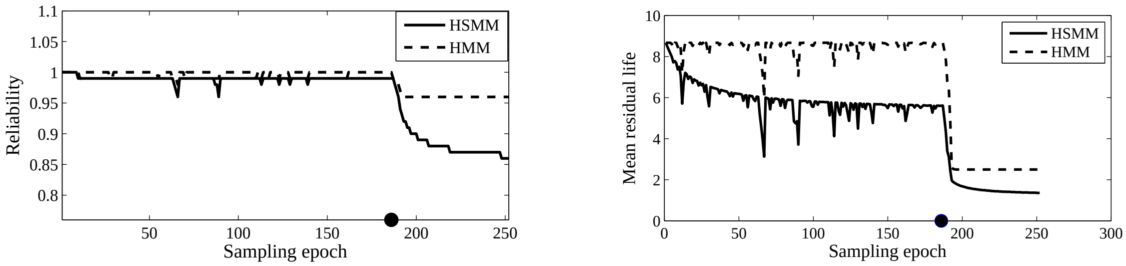
where ARL represents the actual remaining life of a machine at the last sampling epoch and n is the total number of investigated failure histories i.e. $n = 100$. Clearly, lower values of RPD are preferable. RPDs of three different investigated scenarios for both models have been obtained and shown in Table 2. As the results show, the proposed HSMM demonstrates different performances for different shift sizes and we can conclude that the proposed HSMM model outperforms HMM when estimating the remaining residual life in all cases. HSMM produced RPDs which are 49%, 52%, and 66% smaller than those for HMM considering small, medium, and large sizes of the changes. On average, HSMM produced 55.66% lower RPDs in all the three different scenarios when compared with HMM. Therefore, we can confidently assert that HSMM provides more reliable MRL estimates in all three cases investigated.

V. CONCLUSIONS

In this paper, we have applied a hidden semi-Markov modeling for early fault detection and residual life esti-

Table 2: *RPD of the estimated MRL statistic for different values of μ_2*

Model	$\mu_2 = (.8, .6)$	$\mu_2 = (2, 1.5)$	$\mu_2 = (4, 4.5)$
HMM	72%	78%	135%
HSMM	23%	26%	69%



a. Estimated conditional RF   b. Estimated conditional MRL

Figure 2: *Estimated conditional RF and MRL for investigated failure history*

mation for a system subject to deterioration and random failure. The system state process has been modeled as a 3-state hidden semi-Markov process. It has been assumed that vector observations are available at regular sampling times through system condition monitoring. The observations are related to the true underlying state of the system which is unobservable. Given that the sojourn time distributions for both operational states are provided, the unique method for deriving an explicit form of failure time distribution and joint distribution of time to failure and sojourn time have been presented. Using the available failure and suspended data histories obtained from CM, a new parameter estimation procedure has been developed using the EM algorithm and explicit formulas have been derived for the conditional reliability function and the mean residual life of the system. A numerical example has been developed to illustrate the estimation and fault detection procedure. It has been found that both indicators immediately identify the time when the system starts experiencing severe degradation. It has been also shown that the proposed HSMM provides more reliable MRL and RF statistics than HMM which make it suitable for real situations. To compare the effectiveness of the proposed method with other approaches, we plan to develop a case study using real data in future research.

ACKNOWLEDGMENTS

The authors would like to thank the Natural Sciences and Engineering Research Council of Canada and the Ontario Centers of Excellence (OCE) for the financial support under grants number RGPIN 121384-11 and 201461, respectively.

APPENDIX

A.

Let sojourn time distributions in healthy and unhealthy states follow Erlang distribution with probability distribution functions $f_1(t; 2, \lambda_1)$ and $f_2(t; 2, \lambda_2)$, respectively. Since $\mathcal{L}_s(\lambda^2 t e^{-\lambda t}) = \frac{\lambda^2}{(\lambda+s)^2}$ and using Theorem 1 for each $t \in \mathbb{R}^+$, the density function of ξ is given by:

$$\begin{aligned} f_\xi(t) &= p_{12}\mathcal{L}_s^{-1}\Big(\frac{\lambda_1^2}{(\lambda_1+s)^2}.\frac{\lambda_2^2}{(\lambda_2+s)^2}\Big) + p_{13}\lambda_1^2 t e^{-\lambda_1 t} \\ &= p_{12}\lambda_1^2\lambda_2^2\Big(\frac{2e^{-\lambda_1 t}}{(\lambda_1-\lambda_2)^3} - \frac{2e^{-\lambda_2 t}}{(\lambda_1-\lambda_2)^3} \\ &+ \frac{e^{-\lambda_1 t}t}{(\lambda_1-\lambda_2)^2} + \frac{e^{-\lambda_2 t}t}{(\lambda_1-\lambda_2)^2}\Big) + p_{13}\lambda_1^2 t e^{-\lambda_1 t}, \end{aligned}$$

and for all $0 < w < t$,

$$\begin{aligned} & f_{(\xi,\tau_1)}(t,w) = p_{12}\mathcal{L}_{(s,v)}^{-1}\Big(\frac{\lambda_1^2}{(\lambda_1+s+v)^2}.\frac{\lambda_2^2}{(\lambda_2+s)^2}\Big) \\ & + p_{13}\mathcal{L}_{(s,v)}^{-1}\Big(\frac{v}{s+v}\frac{\lambda_1^2}{(\lambda_1+s+v)^2}\Big) \\ & = p_{12}\lambda_1^2\lambda_2^2 e^{-\lambda_2(t-w)-\lambda_1 w}(t-w)wH(t-w) \\ & + p_{13}\lambda_1^2\Big(-e^{-\lambda_1 t}w\delta(-t+w) \\ & - \frac{H(t-w)\delta'(t-w)}{\lambda_1^2} + \frac{e^{-\lambda_1 w}H(t-w)\delta'(t-w)}{\lambda_1^2} \\ & + \frac{e^{-\lambda_1 w}wH(t-w)\delta'(t-w)}{\lambda_1}\Big) \end{aligned}$$

where $H(w-t)$ is a Heaviside step function and $\delta(t)$ is a Dirac Delta function,

$$\delta(t) = \begin{cases} 0 & \text{if} \quad t \neq 0 \\ \infty & \text{if} \quad t = 0 \end{cases}$$

$$H(w-t) = \begin{cases} 0 & \text{if} \quad w < t \\ \frac{1}{2} & \text{if} \quad w = t \\ 1 & \text{if} \quad w > t \end{cases}$$

Since $0 < w < t$, $f_{(\xi,\tau_1)}(t,w)$ can be written as,

$$f_{(\xi,\tau_1)}(t,w) = p_{12}\lambda_1^2\lambda_2^2 e^{-\lambda_2(t-w)-\lambda_1 w}(t-w)w$$

B.

$$Q(\Lambda,\Psi|\hat{\Lambda},\hat{\Psi}) = \sum_{i=1}^{M} Q_{F_i}(\Lambda,\Psi|\hat{\Lambda},\hat{\Psi}) + \sum_{j=1}^{N} Q_{S_j}(\Lambda,\Psi|\hat{\Lambda},\hat{\Psi})$$

For a single failure history,

$$Q_F(\Lambda,\Psi|\hat{\Lambda},\hat{\Psi}) = Q_F^{state}(\Lambda|\hat{\Lambda},\hat{\Psi}) + Q_F^{obs}(\Psi|\hat{\Lambda},\hat{\Psi})$$

The first term Q_F^{state} is given by,

$$Q_F^{state}(\Lambda|\hat{\Lambda},\hat{\Psi}) = \frac{A}{B}$$

where,

$$A = \Big(\int_{w<t} \ln\Big(f_{\tau_1|\xi}(w|t)f_\xi(t)\Big)\,\hat{g}_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,w)\hat{f}_{\tau_1|\xi}(w|t)$$
$$+\ln\Big(m_{\tau_1|\xi}(t|t)f_\xi(t)\Big)\,\hat{g}_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,t)\hat{m}_{\tau_1|\xi}(t|t)\hat{f}_\xi(t)\Big)$$
$$B = \int_{u<t} \hat{g}_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,u)\hat{f}_{\tau_1|\xi}(u|t)f_\xi(t)du$$
$$+\hat{g}_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,t)\hat{m}_{\tau_1|\xi}(t|t)f_\xi(t)$$
$$m_{\tau_1|\xi}(t|t) = 1 - \int_{w<t} f_{\tau_1|\xi}(w|t) = \frac{p_{13}\lambda_1^2 t e^{-\lambda_1 t}}{f_\xi(t)}$$

To simplify the notation, for the remainder of the analysis we denote vectors $\hat{\mathbf{g}} = \big(\hat{g}_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,\Delta),\ldots, \hat{g}_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,T\Delta), \hat{g}_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,t)\big)'$, $\hat{\mathbf{d}}_i = (\hat{d}_i^1,\ldots,\hat{d}_i^T,\hat{d}_i^t)$ for $i = 1,\ldots,5$, and for any vector $\mathbf{v},\mathbf{w}$, $\mathbf{v}'.\mathbf{w} = \langle v,w\rangle$ represents the inner product. Therefore,

$$Q_F^{state}(\Lambda|\hat{\Lambda},\hat{\Psi}) = \frac{\langle\hat{\mathbf{g}},\hat{\mathbf{d}}_1\rangle}{\hat{d}}\ln p_{12} + \hat{\eta}\ln p_{13} + 2(\frac{\langle\hat{\mathbf{g}},\hat{\mathbf{d}}_1\rangle}{\hat{d}} + \hat{\eta})$$
$$\ln\lambda_1 + 2\frac{\langle\hat{\mathbf{g}},\hat{\mathbf{d}}_1\rangle}{\hat{d}}\ln\lambda_2 - \frac{\langle\hat{\mathbf{g}},\hat{\mathbf{d}}_2\rangle}{\hat{d}}\lambda_2 - (\frac{\langle\hat{\mathbf{g}},\hat{\mathbf{d}}_3\rangle}{\hat{d}} + \hat{\eta}t)\lambda_1 + C_1$$

where for $\underline{L} = (k-1)\Delta$ and $\overline{L} = k\Delta$ for $k = 1,\ldots,T$; and $\underline{L} = T\Delta$ and $\overline{L} = t$ for $k = t$,

$$\hat{d}_1^k = \hat{p}_{12}\hat{\lambda_1}^2\hat{\lambda_2}^2 e^{-\hat{\lambda_2}t}\int_{\underline{L}}^{\overline{L}} e^{-w(\hat{\lambda}_1-\hat{\lambda}_2)}w(t-w)dw$$
$$\hat{d}_2^k = \hat{p}_{12}\hat{\lambda_1}^2\hat{\lambda_2}^2 e^{-\hat{\lambda_2}t}\int_{\underline{L}}^{\overline{L}} e^{-w(\hat{\lambda}_1-\hat{\lambda}_2)}w(t-w)^2dw$$
$$\hat{d}_3^k = \hat{p}_{12}\hat{\lambda_1}^2\hat{\lambda_2}^2 e^{-\hat{\lambda_2}t}\int_{\underline{L}}^{\overline{L}} e^{-w(\hat{\lambda}_1-\hat{\lambda}_2)}w^2(t-w)dw$$
$$\hat{d}_4^k = \hat{p}_{12}\hat{\lambda_1}^2\hat{\lambda_2}^2 e^{-\hat{\lambda_2}t}\int_{\underline{L}}^{\overline{L}} e^{-w(\hat{\lambda}_1-\hat{\lambda}_2)}\ln(t-w)w(t-w)dw$$
$$\hat{d}_5^k = \hat{p}_{12}\hat{\lambda_1}^2\hat{\lambda_2}^2 e^{-\hat{\lambda_2}t}\int_{\underline{L}}^{\overline{L}} e^{-w(\hat{\lambda}_1-\hat{\lambda}_2)}\ln(w)w(t-w)dw$$
$$\hat{d} = \langle\hat{\mathbf{d}}_1,\hat{\mathbf{g}}\rangle + \hat{p}_{13}\hat{\lambda_1}^2 e^{-\hat{\lambda_1}t}t\hat{g}_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,t),$$
$$\hat{\eta} = \frac{\hat{p}_{13}\hat{\lambda_1}^2 e^{-\hat{\lambda}_1 t}t\hat{g}_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,t)}{\hat{d}}$$
$$C_1 = \frac{\langle\hat{\mathbf{g}},\hat{\mathbf{d}}_4\rangle}{\hat{d}} + \frac{\langle\hat{\mathbf{g}},\hat{\mathbf{d}}_5\rangle}{\hat{d}} + \hat{\eta}\ln t$$

Also using the same approach,

$$Q_F^{obs}(\Psi|\hat{\Lambda},\hat{\Psi}) = \sum_{k=1}^{T}\hat{c}_k\ln\big(g_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,k\Delta)\big) + \hat{c}_t\ln\big(g_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,t)\big)$$

where,

$$\hat{c}_k = \frac{\hat{d}_1^k\hat{g}_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,k\Delta)}{\hat{d}} \quad \text{for } k = 1,\ldots,T$$
$$\hat{c}_t = \frac{\hat{p}_{13}\hat{\lambda_1}^2 e^{-\hat{\lambda}_1 t}t\hat{g}_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,t)}{\hat{d}}$$

Thus, for $\hat{\mathbf{C}} = (\hat{c}_1,\ldots,\hat{c}_T,\hat{c}_t)$, $\ln\mathbf{g} = \big(\ln g_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,\Delta),\ldots,\ln g_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,T\Delta),\ln g_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,t)\big)'$,

$$Q_F^{obs}(\Psi|\hat{\Lambda},\hat{\Psi}) = \langle\hat{\mathbf{C}},\ln\mathbf{g}\rangle$$

Moreover, following the same approach for a single suspension history, we will have:

$$Q_S(\Lambda,\Psi|\hat{\Lambda},\hat{\Psi}) = Q_S^{obs}(\Psi|\hat{\Lambda},\hat{\Psi}) + Q_S^{state}(\Lambda|\hat{\Lambda},\hat{\Psi})$$

where for $\hat{\mathbf{a}}_i = (\hat{a}_i^1,\ldots,\hat{a}_i^T,\hat{a}_i^t)$ for $i = 1,\ldots,4$ and $\mathbf{a}_5 = (a_5^1,\ldots,\hat{a}_5^T,\hat{a}_5^t)$,

$$Q_S^{state}(\Lambda|\hat{\Lambda},\hat{\Psi}) = \frac{\langle\hat{\mathbf{g}},\hat{\mathbf{a}}_1\rangle}{\hat{f}}\ln p_{12} + 2(\frac{\langle\hat{\mathbf{g}},\hat{\mathbf{a}}_1\rangle}{\hat{f}} + \hat{v})\ln\lambda_1$$
$$-\frac{\langle\hat{\mathbf{g}},\hat{\mathbf{a}}_2\rangle}{\hat{f}}\lambda_2 + \frac{\langle\hat{\mathbf{g}},\mathbf{a}_5\rangle}{\hat{f}} - (\frac{\langle\hat{\mathbf{g}},\hat{\mathbf{a}}_3\rangle}{\hat{f}} + \hat{q})\lambda_1 + \mathrm{C}_2$$

where,

$$\hat{a}_1^k = \hat{p}_{12}\hat{\lambda_1}^2 e^{-\hat{\lambda}_2 t}\int_{\underline{L}}^{\overline{L}} we^{-\hat{\lambda_1}w+\hat{\lambda_2}w}(1+\hat{\lambda}_2(t-w))dw$$

$$\hat{a}_2^k = \hat{p}_{12}\hat{\lambda_1}^2 e^{-\hat{\lambda}_2 t}\int_{\underline{L}}^{\overline{L}} w(t-w)e^{-\hat{\lambda_1}w+\hat{\lambda_2}w}(1+\hat{\lambda}_2(t-w))dw$$

$$\hat{a}_3^k = \hat{p}_{12}\hat{\lambda_1}^2 e^{-\hat{\lambda}_2 t}\int_{\underline{L}}^{\overline{L}} w^2 e^{-\hat{\lambda_1}w+\hat{\lambda_2}w}(1+\hat{\lambda}_2(t-w))dw$$

$$\hat{a}_4^k = \hat{p}_{12}\hat{\lambda_1}^2 e^{-\hat{\lambda}_2 t}\int_{\underline{L}}^{\overline{L}} w\ln(w)(1+\hat{\lambda}_2(t-w))e^{-\hat{\lambda_1}w+\hat{\lambda_2}w}dw$$

$$a_5^k = \hat{p}_{12}\hat{\lambda_1}^2 e^{-\hat{\lambda}_2 t}\int_{\underline{L}}^{\overline{L}} w\ln(1+\lambda_2(t-w))e^{-\hat{\lambda_1}w+\hat{\lambda_2}w}$$
$$(1+\hat{\lambda}_2(t-w))dw$$

$$\hat{f} = \langle \hat{\mathbf{a}}_1, \hat{\mathbf{g}}\rangle + \hat{g}_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,t)\hat{\lambda_1}^2\int_t^\infty ue^{-\hat{\lambda_1}u}du;$$

$$\hat{q} = \frac{\hat{\lambda}_1^2 \hat{g}_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,t)}{\hat{f}}\int_{w>t} w^2 e^{-\hat{\lambda}_1 w}dw$$

$$C_2 = \frac{\langle \hat{\mathbf{g}}, \hat{\mathbf{a}}_4\rangle}{\hat{f}} + \hat{\lambda}^2 \hat{g}_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,t)\int_t^\infty w\ln(w)e^{-\hat{\lambda}_1 w}dw;$$

$$\hat{v} = \frac{\hat{\lambda}_1^2 \hat{g}_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,t)}{\hat{f}}\int_{w>t} we^{-\hat{\lambda}_1 w}dw$$

and also,

$$Q_S^{obs}(\Psi|\hat{\Lambda},\hat{\Psi}) = \langle \hat{\mathbf{D}}, \ln \mathbf{g}\rangle$$

where $\hat{\mathbf{D}} = (\hat{D}^1,\ldots,\hat{D}^T, D^t)$, and

$$\hat{D}^k = \frac{\hat{a}_1^k \hat{g}_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,k\Delta)}{\hat{f}} \quad \text{for } k = 1,\ldots,T$$

$$\hat{D}^t = \frac{\hat{g}_{\vec{Y}|\xi,\tau_1}(\vec{y}|t,t)\hat{\lambda_1}^2\int_t^\infty ue^{-\hat{\lambda_1}u}du}{\hat{f}}$$

References

[1] M Dong and D He. Hidden semi-markov model-based methodology for multi-sensor equipment health diagnosis and prognosis. *European Journal of Operational Research*, 178(2):858–878, 2007.

[2] A K S Jardine, D Lin, and D Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20:1483–1510, 2006.

[3] R Jiang, M J Kim, and V Makis. Maximum likelihood estimation for a hidden semi-markov model with multivariate observation. *Quality and Reliability Engineering International*, 2012.

[4] M J Kim and V Makis. Joint optimization of sampling and control of partially observable failing systems. *Operations Research*, 61(3):77–790, 2013.

[5] M J Kim, V Makis, and R Jiang. Parameter estimation for partially observable systems subject to random failure. *Applied Stochastic Models in Business and Industry*, (279–294), 2012.

[6] V Makis. Multivariate bayesian control chart. *Operations Research*, 56:487–496, 2008.

[7] V Makis, J Wu, and Y Gao. An application of dpca to oil data for cbm modeling. *European Journal of Operational Research*, 174:112–123, 2006.

[8] R Moghaddass and M Zuo. Multistate degradation and supervised estimation methods for a condition-monitored device. *IIE Transactions*, 46:131–148, 2014.

[9] Y Peng and M Dong. A prognosis method using age-dependent hidden semi-markov model for equipment health prediction. *Mechanical Systems and Signal Processing*, 25:237–252, 2011.

[10] X-S. Si, W. Wangb, C-H. Hua, and D-H. Zhou. Remaining useful life estimation: A review on the statistical data driven approaches. *European Journal of Operational Research*, 213:1–14, 2011.

[11] W Wang and AH Christer. A general condition based maintenance model for a stochastic dynamic system. *Journal of the Operational Research Society*, 51:145–155, 2000.

[12] X Wang, V Makis, and M Yang. A wavelet approach to fault diagnosis of agearbox under varying load conditions. *Journal of Soundand Vibration*, 329:1570–1585, 2010.

[13] J Yang and V Makis. Dynamic response of residual to external deviations in a controlled production process. *Technometrics*, 42:290–299, 2000.

[14] Z Zhou, C Hu, D Xu, M Chen M, and D Zhou. A model for real-time failure prognosis based on hidden markov-model and belief rule base. *European Journal of Operational Researh*, 207:269–283, 2010.