Condition Monitoring and Remaining Useful Life Prediction Using
Degradation Signals: Revisited

Nan Chen

Department of Industrial and Systems Engineering, National University of Singapore

Kwok Leung Tsui

Department of Systems Engineering and Engineering Management, City University of Hong Kong

June 20, 2012

Abstract

Condition monitoring is an important prognostic tool to determine the current operation status of the system/device, and to estimate the distribution of the remaining useful life. In this paper, we propose a two-phase model to characterize the degradation process of the rotational bearings. We use the Bayesian framework to integrate historical data with up-to-date in-situ observations of new working units to improve the degradation modeling and prediction. We also develop a new approach to compute the distribution of the remaining useful life based on the degradation signals, which is more accurate compared with that in the literature. At last, we present extensive numerical results to demonstrate that our new framework is effective and efficient.

Keywords: Condition monitoring, Degradation, Remaining useful life, Bayesian
1 Introduction

Modern engineering systems are overwhelmingly complex because of the increasing requirement on their functionalities and qualities. These systems often have high standard on system reliability because a single failure can lead to catastrophic consequences with profound impacts, extreme costs, and potential safety hazards. Therefore, effective methods that can predict and prevent system failures have long been sought after. Unfortunately, traditional ways in predicting system failures often fail in practice (Pecht 2008) because the knowledge of failure mechanisms is limited or even unknown in complex systems.

At the same time, the fast development of information and sensing technologies provides us with tremendous opportunities to develop a new set of methodologies to maintain a reliable and healthy system. The in-situ sensing data are often collected and analyzed during system operations to characterize and predict systems' health conditions (Nelson 1990), to prepare necessary preventive maintenance (e.g. Wang et al. 2012; Wang 2007; Chen et al. 2011), or even to plan burn-in for future products (Ye et al. 2012a,b). Built on these real time sensing data, condition monitoring methods use statistical models to emulate the physical degradation processes. They have been proved to be more flexible and widely applicable compared with physical models which rely on thorough knowledge of the failure mechanisms. Methods in condition monitoring can be classified into two categories depending on whether the health state of the systems are directly observable or not. In the first category where the health state is not directly observable, different models (e.g. Baruah and Chinnam 2005; Kumar and Klefsj 1994) are used to link other observable signals or environmental/operational factors with the unobservable health state to predict the remaining useful life (RUL). In the second category where the health state is observable, evolutions of the health condition are often directly characterized by random coefficient models, Brownian motions, Gamma processes, etc (e.g. Doksum and Hyland 1992; Lu and Meeker 1993; Park and Padgett 2005; Ye et al. 2012c; Si et al. 2012). Si et al. (2011) provided an excellent reviews of both categories of methods to predict the RUL using condition monitoring data.

In this paper, we investigate the condition monitoring and RUL prediction of the rotational bearings. The bearing degradation is manifested by the magnitude of vibration during rotations. After the bearing being installed, it will experience an initial stable stage, when the vibration is
slight. But its degradation becomes more and more severe after an unknown change point, when the vibration magnitude increases dramatically and features large variability. The bearing is considered to have failed once its degradation exceeds certain predetermined threshold. However, different bearings often have different locations of change points and large variability in the increment rates of vibration magnitude, despite the similar shapes of their degradation paths, as illustrated in Figure 1. Gebraeel et al. (2005), among other notable works, proposed a Bayesian framework to model this type of degradation signals. They used an exponential model with random coefficients to characterize the evolution of degradation signals after the change point

\[ L_j \equiv \log[S(t_j) - \delta] = a + b \cdot t_j + \epsilon_j, \]

where \( S(t_j) \) is the vibration magnitude observed at time \( t_j \), \( \delta \) is a known constant, \( a, b \) are the coefficients (parameters) of the models, and \( \epsilon_j \) is the corresponding error term. Under the Bayesian framework they have developed, \( a, b \) are random variables following certain prior distributions, and \( \epsilon_j \) follows i.i.d. normal distribution. When the vibration signals are observed, the posterior distribution of \( a, b \) can be computed. The model (1) with parameters following the posterior distribution \( P(a,b|L_1, L_2, \cdots, L_n) \) is expected to more accurately characterize the degradation of
the device, from which $L_1, L_2, \ldots, L_n$ are observed. The Bayesian method they used leads to a simple yet effective way to fuse the information between historical data and sensing signals of the current working unit. Other representative works along the similar line can be found in (Gebraeel 2006; Chakraborty et al. 2009).

Despite their importance, there are several aspects that can be improved. First, existing works often start condition monitoring after the change point of the degradation as if they know the exact location of the change point. However, the change points usually vary from one unit to another significantly, as clearly illustrated in Figure 1. By considering the distribution of the change point, we may be able to predict the RUL right after the installment. Second, the variance of $\epsilon_j$ is often assumed to be the same among different units. However, it is not uncommon in practice that units are heterogeneous, leading to different variations in the degradation paths, as demonstrated in Figure 1 as well. Consequently, appropriate modeling of the variance heterogeneity is also expected to improve the prediction accuracy. Third and most importantly, in the Bayesian framework, the predictions of degradation at different future times are correlated even if the past observations are i.i.d. because they are predicted using the same posterior distribution of the model parameters. If the correlation among future predictions are not considered appropriately, the distribution of RUL might be inaccurate. In this paper, we improve these aspects in modeling bearing degradations. We propose a two-phase threshold model to explicitly account for the different phases of the degradation. When new observations are available, we update the posterior distributions of model parameters, including regression coefficients and the variance of the error term using Bayesian methods. We also propose a new approach, which takes the correlations among degradation predictions into consideration, to compute the RUL distribution with better accuracy. We would like to stress that our approach can be naturally extended to more general degradation models. For example, it can be applied when the degradation path is nonlinear, by choosing appropriate basis functions in the linear model (De Boor 2001).

The rest of the paper is organized as follows. Section 2 presents the degradation model, and illustrates the overall framework of the condition monitoring. Section 3 discusses how to specify the prior distributions of the model parameters. Section 4 presents the technical details on how to update the model parameters and predict the RUL of a new working unit using Bayesian method. Section 5 demonstrates the effectiveness of the proposed method using numerical simulations and
examples from real dataset. And Section 6 concludes the paper with discussions on future research.

2 Two-phase Degradation Model

In this paper, we propose a two-phase degradation model to explicitly account for the change point of the degradation. Models with change points are commonly used when there are distinguished phases in the degradation signals (Bae and Kvam 2006; Ng 2008), or hazard rates (Lin 2008; Loader 1991; Yuan and Kuo 2010). In our approach, we adopted the similar piecewise log-linear model as used by Bae and Kvam (2006), but we use the Bayesian approach to update the model parameters and make prediction.

We denote \( L_{ij} \) as the (transformed) degradation signal of unit \( i \) observed at time \( t_{ij} \). We can express the two-phase degradation model as

\[
L_{ij} = \ln [S(t_{ij}) - \delta] = \begin{cases} 
  a_{i1} + b_{i1}t_{ij} + \sigma_{i1}\epsilon_{ij}, & t_{ij} \leq \gamma_i \\
  a_{i2} + b_{i2}(t_{ij} - \gamma_i) + \sigma_{i2}\epsilon_{ij}, & t_{ij} > \gamma_i 
\end{cases}
\]

where \( \epsilon_{ij} \) follows i.i.d. standard normal distribution; \( a_{i1}, a_{i2}, b_{i1}, b_{i2}, \sigma_{i1}, \sigma_{i2} \) are the parameters in each of the two phases in the degradation of unit \( i \). \( \gamma_i \) is the random variable denoting the change point of unit \( i \). By setting \( \gamma_i = 0 \), and \( \sigma_{i2} \) the same for all \( i \), our model reduces to the simpler model considered in (Gebraeel et al. 2005; Gebraeel 2006). From Figure 1, we can find that in the degradation of rotational bearings, usually \( b_{i1} \) is close to zero, and \( \sigma_{i1} \ll \sigma_{i2} \). In the following, we denote \( \beta_{i1} = [a_{i1}, b_{i1}]^T, \beta_{i2} = [a_{i2}, b_{i2}]^T, \theta_i = [\beta_{i1}^T, \sigma_{i1}, \beta_{i2}^T, \sigma_{i2}]^T \) for notation simplicity.

To improve the modeling accuracy, we use the Bayesian approach to integrate the historical data with sensing observations of a working unit. Historical degradations of failed units provide the information on the possible values of the model parameters \( \theta_i, \gamma_i \) of the same type of units. And the in-situ sensing observations make the model parameters more and more accurate toward the true value of the particular working unit. In more details, we assume the model parameters \( \theta_i, \gamma_i \) are random variables following a common prior distribution with density \( \pi(\theta_i, \gamma_i) \), which can be learnt from historical data. When sensing observations from a new working unit \( i \) are available up to current time \( \tau \), we can update the posterior distribution of the parameters \( \theta_i, \gamma_i \) using the
Historical data from other bearings

Vibration signals up to time \( t \)

Estimating prior distributions

Updated model of degradation

Prediction of residual life

Condition based maintenance

new observations

Condition monitoring

Bayes formula (Gelman et al. 2004)

\[
P(\theta_i, \gamma_i | L_i(\tau)) \propto P(L_i(\tau)|\theta_i, \gamma_i) \cdot \pi(\theta_i, \gamma_i),
\]

where \( L_i(\tau) \equiv [L_{i1}, L_{i2}, \cdots, L_{ij}], \forall t_{ij} \leq \tau \) are the observations of the degradation magnitude, and \( P(L_i(\tau)|\theta_i, \gamma_i) \) is the likelihood function based on the model (2). With the sensing information \( L_i(\tau) \), the model with updated parameters would fit the degradation of unit \( i \) better. In addition, to compute the RUL, we can also predict future degradations \( L_{ik} \), \( \forall t_{ik} > \tau \) of the unit \( i \) based on its updated degradation model

\[
P(L_{ik}|L_i(\tau)) = \int P(L_{ik}|\theta_i, \gamma_i, L_i(\tau)) \cdot P(\theta_i, \gamma_i|L_i(\tau)) d\theta_i d\gamma_i.
\]

The overall work flow of condition monitoring, including model updating and RUL prediction, is illustrated in Figure 2.

Despite the simple and straightforward formulation, direct computations of (3) and (4) are time consuming because they involve multi-dimensional integrations, and require long Monte Carlo simulation time to achieve the acceptable accuracy. Therefore, they are prohibitive in real time condition monitoring. Clearly we need a less computational demanding way to update the degradation model and predict RUL efficiently. Targeting on this gap, we propose an empirical Bayes approach to mitigate the computation load through proper selection of the prior distributions and
efficient strategy of the model updating.

3 Specification of Priors

Prior information is an unique and important component in Bayesian framework. It offers the opportunity to integrate the domain knowledge or past experience with the newly observed data effectively. Accurate and informative priors can improve the accuracy of the degradation model when observations of the working unit are limited. In practice, priors can be derived from domain knowledge and expert experience, or be estimated from historical data. In this paper, we take the objective approach by estimating the priors from historical degradation data of the failed units, as did by Gebraeel (2006).

Assuming the prior distribution can be expressed as some parametric model $\pi(\theta, \gamma | \zeta)$ with hyper-parameters $\zeta$, the empirical Bayes method can be adopted to estimate these hyper-parameters by maximizing the marginal likelihood

$$
\hat{\zeta} = \arg \max_{\zeta} \prod_{i=1}^{I} \int P(L_i | \theta_i, \gamma_i) \cdot \pi(\theta_i, \gamma_i | \zeta) d\theta_i d\gamma_i,
$$

where $L_i = [L_{i1}, L_{i2}, \cdots, L_{ini}]^T$ is the historical degradations of unit $i$, $n_i$ is the number of degradation observations, and $\theta_i, \gamma_i$ are the parameters of the degradation model of unit $i$. Unfortunately, if the model is complex or if the parametric form of the prior distribution is incorrect, this method may not perform well. In this paper, we propose an alternative approach to estimate the prior distributions. Instead of maximizing the marginal distribution of the historical data, we consider the maximum likelihood estimates (MLE) of their model parameters $\hat{\theta}_i, \hat{\gamma}_i$ of each unit for $i = 1, 2, \cdots, I$, as samples from the prior distributions. Subsequently, we can estimate the prior distributions using distribution selection and fitting techniques such as histograms or $\chi^2$ goodness of fit tests.

Compared with the model in (Gebraeel et al. 2005), the MLE in our model is more complicated due to the discontinuity introduced by the change point parameter $\gamma_i$. By definition, the log
likelihood function given observations $L_i$ of unit $i$ can be written as

$$\begin{align*}
    l(\theta_i, \gamma_i|L_i) &= \sum_{j=1}^{n_i} \left[ -\frac{1}{2} \ln(2\pi\sigma_{i1}^2) - \frac{(L_{ij} - a_{i1} - b_{i1}t_{ij})^2}{2\sigma_{i1}^2} \right] \cdot I(t_{ij} \leq \gamma_i) \\
    &\quad + \sum_{j=1}^{n_i} \left[ -\frac{1}{2} \ln(2\pi\sigma_{i2}^2) - \frac{(L_{ij} - a_{i2} - b_{i2}t_{ij} + b_{i2}\gamma_i)^2}{2\sigma_{i2}^2} \right] \cdot I(t_{ij} > \gamma_i),
\end{align*}$$

(5)

where $I(\cdot)$ is the indicator function, which equals one when the condition is true and zero otherwise. It is difficult to maximize (5) directly due to the discontinuities. Instead, we can consider the likelihood conditioned on $\gamma_i$, i.e., $l(\theta_i|L_i, \gamma_i)$. With fixed $\gamma_i$, the parameters $a_{i1}, b_{i1}, \sigma_{i1}$ and $a_{i2}, b_{i2}, \sigma_{i2}$ are well separated in the likelihood function, and can be estimated using standard linear regression results (Seber and Lee 2003). We can get the closed form solution of $\beta_{im}$ and $\sigma_{im}$ for $m = 1, 2$ that maximizes the $l(\theta_i|L_i, \gamma_i)$ as:

$$\hat{\beta}_{im} = (X_{im}^T X_{im})^{-1}X_{im}^T Y_{im}, \quad \hat{\sigma}^2_{im} = \frac{1}{n_{im}}(Y_{im} - X_{im}\hat{\beta}_{im})^T(Y_{im} - X_{im}\hat{\beta}_{im}), \quad m = 1, 2$$

(6)

where $n_{i1} = \sum_{j=1}^{n_i} I(t_{ij} \leq \gamma_i)$, and $n_{i2} = n_i - n_{i1}$ are the number of observations in each degradation phase respectively; and the matrices are defined as

$$Y_{i1} = [L_{i1}, L_{i2}, \cdots, L_{i[\gamma_i]}]^T, \quad Y_{i2} = [L_{i[\gamma_i]}, \cdots, L_{im_i}]^T$$

(7)

$$X_{i1} = \begin{bmatrix}
    1, & 1, & \cdots, & 1 \\
    t_{i1}, & t_{i2}, & \cdots, & t_{i[\gamma_i]}
\end{bmatrix}^T, \quad X_{i2} = \begin{bmatrix}
    1, & \cdots, & 1, & 1 \\
    t_{i[\gamma_i]} - \gamma_i, & \cdots, & t_{i(n_i-1)} - \gamma_i, & t_{im_i} - \gamma_i
\end{bmatrix}^T$$

where $[\gamma_i]$ ($[\gamma_i]$) are the largest (smallest) integer such that $t_{i[\gamma_i]} \leq \gamma_i$ ($t_{i[\gamma_i]} > \gamma_i$). It is worth mentioning here that the quantities $\hat{\beta}_{im}, \hat{\sigma}^2_{im}, n_{im}$ are functions of $\gamma_i$. Here we do not explicitly express this dependence to avoid tedious notations. The conditional maximum log-likelihood, which is defined as $M(\gamma_i) = \max_{\theta_i} l(\theta_i|L_i, \gamma_i)$, can therefore be expressed as

$$M(\gamma_i) = -\frac{1}{2} \sum_{m=1}^{2} \left[ \frac{n_{im}}{2} + \frac{n_{im}}{2} \cdot \ln 2\pi \|Y_{im} - X_{im}(X_{im}^T X_{im})^{-1}X_{im}^T Y_{im}\|^2 \right],$$

(8)

where $\| \cdot \|^2$ denotes the $L_2$ norm of a vector. It can be proved that $\hat{\gamma}_i = \arg \max M(\gamma_i)$ together with $\hat{\theta}_i = \arg \max l(\theta_i|L_i, \hat{\gamma}_i)$ jointly maximize the likelihood function in (5). From the derivation,
we can find that $X_{im}, Y_{im}$ only changes when $\gamma_i$ crosses some observation point $t_{ij}$. Therefore, $M(\gamma_i)$ is a stepwise constant function with at most $n_i + 1$ different values. Hence, $\gamma_i$ exists and is finite. We would like to point out that there also exist other methods in determining the location of $\gamma_i$, such as (Lai 1995; Siegmund and Venkatraman 1995). But to limit the scope of the paper, we do not discuss alternative methods here further.

We can use the maximum likelihood estimates $(\hat{\theta}_i, \hat{\gamma}_i)$ for all the units $i = 1, 2, \cdots, I$ failed in the past to estimate the prior distribution. In our model, the prior distribution is multidimensional, and requires very large number of samples to estimate without any constraints. Given the limited number of historical samples, we need to simplify the prior distribution by assuming the conditional independence of the model parameters between two phases

$$
\pi(\theta_i, \gamma_i) = \pi(\gamma_i)\pi(\beta_{i1}, \sigma_{i1}|\gamma_i)\pi(\beta_{i2}, \sigma_{i2}|\gamma_i).
$$

Additionally, for certain prior distributions, or so called conjugate priors, the posterior distributions can be readily obtained without numerical integration. This feature is extremely important in condition monitoring since timely update of degradation model is highly desired. Therefore, in this paper, we also prefer using conjugate priors for efficient model updating. Based on our model, we can specify the priors as

$$
\pi(\beta_{im}, \sigma_{im}|\gamma_i) = \pi(\beta_{im}|\sigma_{im}^2, \gamma_i)\pi(\sigma_{im}^2|\gamma_i) = N(\mu_m, \sigma_m^2\Sigma_m) \cdot SI\chi^2(v_m, s_m^2), \quad m = 1, 2
$$

where $N(\mu_m, \sigma_m^2\Sigma_m)$ is normal distribution with density function

$$
f(\beta; \mu_m, \sigma_m^2\Sigma_m) = \frac{1}{(2\pi)^{\kappa/2}\sigma_m^{\kappa/2}(\det \Sigma_m)^{1/2}} \cdot \exp \left[ -\frac{(\beta - \mu_m)^T\Sigma_m^{-1}(\beta - \mu_m)}{2\sigma_m^2} \right],
$$

where $\kappa$ is the dimension of $\beta$. $SI\chi^2(v_m, s_m^2)$ is the scaled inverse $\chi^2$ distribution with density

$$
f(\sigma^2; v_m, s_m^2) = \frac{(v_ms_m^2/2)^{v_m/2}}{\Gamma(v_m/2)}(\sigma^2)^{-(v_m/2+1)} \exp \left[ -\frac{v_m s_m^2}{2\sigma^2} \right].
$$

Here $\Gamma(\cdot)$ is the Gamma function, and $\mu_m, \Sigma_m, v_m, s_m^2, \quad m = 1, 2$ are the hyper-parameters need to be estimated to specify the prior distribution $\pi(\beta_{im}, \sigma_{im}|\gamma_i)$. We can estimate these hyper-
parameters using MLE based on the samples \((\hat{\theta}_i, \hat{\gamma}_i), i = 1, 2, \cdots, I\) from historical data.

Compared with the distribution \(\pi(\beta_{im}, \sigma_{im}|\gamma_i), \pi(\gamma_i)\) can be more flexible. Possible candidates include uniform distribution, normal distribution, exponential distribution, etc., and can be determined from the historical data by certain goodness-of-fit tests.

4 Bayesian Approach for Updating and Prediction

In this section, we will discuss how to update the model parameters and predict the RUL of a new working unit at any time point based on its sensing observations. If we assume different units work independently, their updating and prediction procedures are the same and can be done independently. Therefore, we will omit the unit index \(i\) hereafter in this section. In addition, we will consider the updating and prediction at a single time \(\tau\). But this procedure can be repeated at multiple different times when new observations are available.

4.1 Parameter updating

When the unit is in operation, and its degradation signals have been observed periodically until current time \(\tau\), we can update the degradation model of the unit by computing the posterior distribution of its parameter \(\theta, \gamma\). This model updating can be performed at any time, and should be done regularly as more sensing observations become available.

We denote the degradation magnitudes of the working unit at time \(0 \leq t_1 < t_2 < \cdots < t_n < \tau\) as \(L = [L_1, L_2, \cdots, L_n]^T\). Using Bayes formula we have

\[
P(\theta, \gamma|L) \propto P(L|\theta, \gamma) \cdot \pi(\theta, \gamma).
\]

However, computing this posterior distribution is time consuming, which is not suitable in the real time condition monitoring. To reduce the computational load, we employ a two-step empirical Bayesian method (Carlin and Louis 2000). Specifically, in the first step, we identify the most probable location of the change point or the most probable region it locates using the posterior mode \(\hat{\gamma}\) of \(P(\gamma|L)\). In the second step, we consider \(\hat{\gamma}\) as the real change point, and perform the updating of \(\theta\) using \(P(\theta|L, \hat{\gamma})\). In practice, as the number of observations increases, \(P(\gamma|L)\) concentrates more
Updating at $\tau = 150$

Updating at $\tau = 450$

Figure 3: Posterior probability of the location of change point at different time The upper panels are the degradation observations up to two time points, and the bottom panels are the corresponding posterior probability $P(\gamma | L)$.

and more around $\hat{\gamma}$. Therefore, this approximation becomes more accurate in subsequent updates. Figure 3 illustrates the posterior distribution $P(\gamma | L)$ at two different updating times. In the left part, the model is updated when the degradation is still in the first phase. The posterior mode equals to the last observation time $t_n$, which indicates that no change point is suspected in the previous observations. In the right part, the model is updated when the degradation has already entered the second phase. And the posterior mode finds the real change point correctly.

After $\hat{\gamma}$ has been identified, we can further compute the posterior distribution of $\theta$. Instead of integrating out $\gamma$ given the unit’s degradation observations $L$, we use a point mass distribution at $\hat{\gamma}$ to approximate the conditional distribution $P(\gamma | L)$, i.e.,

$$P(\theta | L) = \int P(\theta, \gamma | L) d\gamma = \int \frac{P(L|\theta, \gamma) \cdot \pi(\theta | \gamma) \cdot \pi(\gamma)}{P(L)} d\gamma = \int \frac{P(L|\theta, \gamma) \cdot \pi(\theta | \gamma)}{P(L|\gamma)} \cdot P(\gamma | L) d\gamma \approx \frac{P(L|\theta, \hat{\gamma}) \cdot \pi(\theta | \hat{\gamma})}{P(L|\hat{\gamma})}.$$ (10)

The last approximation in (10) is valid because $P(\gamma | L)$ concentrates around $\hat{\gamma}$ with a reasonable sample size, as shown in Figure 3. Using this empirical Bayes approximation, the posterior dis-
Theorem 1 Given $\hat{\gamma}$, the posterior distribution $P(\beta_1, \sigma_1|L, \hat{\gamma})$ and $P(\beta_2, \sigma_2|L, \hat{\gamma})$ are independent. And for $m = 1, 2$, $P(\sigma_m^2|L, \hat{\gamma})$ follows scaled inverse $\chi^2$ distribution with parameter $\hat{v}_m = v_m + n_m$ and
\[
\hat{s}_m^2 = \left[ v_m s_m^2 + Y_m^T Y_m + \mu^T \Sigma^{-1}_m \mu - \hat{\mu}_m^T (X_m^T X_m + \Sigma^{-1}_m) \hat{\mu}_m \right] / (v_m + n_m).
\]
And $P(\beta_m|\sigma_m^2, L)$ follows normal distribution with mean $\hat{\mu}_m = (X_m^T X_m + \Sigma^{-1}_m)^{-1} (X_m^T Y_m + \Sigma^{-1}_m \mu)$, and variance $\sigma_m^2 (X_m^T X_m + \Sigma^{-1}_m)^{-1}$. The matrices $X_m, Y_m$ are determined through (7) with $\gamma_i$ substituted by $\hat{\gamma}$.

From the procedure, we can find that the overall updating rules only need some matrix operations, which are efficient and suitable for real time updating. In practice, we can update the model upon arrival of every new observation or a group of observations depending on the sampling frequency of the data as well as the variability of the data. We would like to point out when $\hat{\gamma}$ equals to the last observation time, i.e., $\hat{\gamma} = t_n$, we have $Y_1 = L, Y_2 = 0$. Therefore, it is not necessary to update the parameters in the second phase since no observations from there are available. However, it does not imply the real change point is likely to be around $t_n$. In fact, it can only indicate that the real change point is likely to be larger than $t_n$. In other words, $P(\gamma|L)$ can not be approximated by a degenerated point mass distribution at $\hat{\gamma}$. It might be approximated by the conditional distribution $\pi(\gamma|\gamma > \hat{\gamma})$ instead. Although this difference will not influence the updating rule of $\theta$, it will change the prediction of RUL, as discussed in Section 4.2.

### 4.2 Remaining useful life prediction

After the degradation model of the working unit being updated using the available sensing information, we can predict the degradation magnitude at any specified time in the future. Since the updated model integrates both prior information and its own specific degradation feature, it is expected to provide more accurate and “customized” predictions. Unlike the degradation model where the variance of $\epsilon$ is treated as a constant, the distribution of the predicted degradation is no longer normal. Additionally, the two-phase model adds another dimension of complexity in RUL prediction.
In the model updating, there are two scenarios depending on the estimated location of the change point $\hat{\gamma}$. These two scenarios have different prediction procedures. First we consider the case when $\hat{\gamma} < t_n$, which means the degradation has been in the second phase at time $t_n$. If the empirical Bayes approximation is valid, in other words $\hat{\gamma}$ is close to the real change point, the prediction only depends on $\gamma, \beta_2, \sigma_2$ according to our degradation model (2). Consequently the predicted degradation magnitudes at a set of future observation times can be computed, based on which the distribution of RUL can be obtained. Specifically, if we assume that the unit can only fail over an arbitrary countable set of time points (e.g., the inspection times, or the periodical sensing times), we can compute the RUL distribution analytically, as stated in Theorem 2 (the proof is included in the Appendix A.2). It is worth noticing that this assumption is not restrictive in practice, and is imposed only to avoid mathematical difficulties. In fact, when the interval between neighboring time points is small, it would be accurate enough from practical point of view even if the failure time is continuously distributed.

**Theorem 2**

Given the collection of possible failure times after current time $\tau$, $\mathcal{T}_\tau = \{T_k|k = 1, 2, \cdots, T_k > \tau \geq \hat{\gamma}\}$, the RUL distribution $P(R_\tau \leq T_k - \tau|L) = 1 - MT_k(K)$ where $R_\tau$ denotes the RUL at time $\tau$; $K$ is $k$ dimensional vector with each entry $K$ denoting the failure threshold; $MT_k(\cdot)$ is the CDF of a $k$ dimensional multivariate $t$ distribution with degree of freedom $\hat{\nu}_2$, mean $\bar{X} \hat{\mu}_2$ and squared scale matrix $\hat{s}_2^2[I + \bar{X}(X_2^T X_2 + \Sigma_2^{-1})^{-1} \bar{X}^T]$. $\bar{X}$ is the matrix defined as

$$
\bar{X}^T = \begin{bmatrix} 1, 1, \cdots, 1 \\
T_1 - \hat{\gamma}, T_2 - \hat{\gamma}, \cdots, T_k - \hat{\gamma} \end{bmatrix}_{2 \times k}.
$$

Theorem 2 provides a way to compute the distribution of RUL at any given time point $\tau$. One distinct feature here is that the correlations among predicted values at different time points are considered even though the the observations are assumed independent. This feature improves the prediction accuracy of RUL compared with traditional methods, as demonstrated in Section 5.

In the second scenario, i.e. $\hat{\gamma} = t_n$, the prediction is more complicated since we need to consider the possible locations of the unknown change point. As discussed in Section 4.1, we can use $\pi(\gamma|\gamma > \hat{\gamma})$ to approximate the posterior distribution $P(\gamma|L)$. First we consider the prediction
of the degradation magnitude $L_{T_k}$ at a single future time point $T_k > \tau$. Depending on whether $T_k$ is before or after the (unknown) change point, we need to predict using different phases of degradation model:

$$P(L_{T_k}|L) = \int_{T_k}^{\infty} P(L_{T_k}|\beta_1, \sigma_1) \cdot P(\beta_1, \sigma_1|L, \gamma) \cdot \pi(\gamma|\gamma > \tilde{\gamma}) d\gamma$$

$$+ \int_{\tilde{\gamma}}^{T_k} P(L_{T_k}|\beta_2, \sigma_2) \cdot P(\beta_2, \sigma_2|L, \gamma) \cdot \pi(\gamma|\gamma > \tilde{\gamma}) d\gamma. \quad (11)$$

Since the observations $L$ are only available up to $t_n = \tilde{\gamma}$, the posterior distributions of model parameters $\theta$ will be the same given $L$ regardless the true change point $\gamma$. Consequently, according to the similar derivation in the proof of Theorem 2, (11) becomes a mixture of $t$-distribution with weight $\omega_1 = \int_{T_k}^{\infty} \pi(\gamma|\gamma > \tilde{\gamma}) d\gamma$ for the prediction based on the first phase and $\omega_2 = 1 - \omega_1$ for the second phase respectively. However, it becomes intractable to find the joint distribution of the degradation magnitudes at multiple future time points, though the similar reasonings still apply.

In practice, as demonstrated in Figure 1, the degradation can hardly cross the failure threshold when the unit is still in the first phase. Equivalently, we can say that the failure time is larger than $\gamma$ almost certainly. Therefore the RUL over the set of possible failure times after $\tau$, $\mathcal{F}_\tau = \{T_k|k = 1, 2, \cdots, T_k > \tau\}$, can be computed as

$$P(R_\tau > T_k - \tau|L) = P(L_{T_1} \leq K, L_{T_2} \leq K, \cdots, L_{T_k} \leq K|L)$$

$$\approx \sum_{s=1}^{k} \int_{T_{s-1}}^{T_s} P(L_{T_s} \leq K, L_{T_{s+1}} \leq K, \cdots, L_{T_k} \leq K|L, \gamma) \cdot \pi(\gamma|\gamma > \tilde{\gamma}) d\gamma$$

$$+ \int_{T_k}^{\infty} \pi(\gamma|\gamma > \tilde{\gamma}) d\gamma, \quad (12)$$

where we define $T_0 = \tilde{\gamma}$ for notation simplicity, and use $L_{T_s}$ to denote the degradation at time $T_s$. We can compute $P(L_{T_s} \leq K, L_{T_{s+1}} \leq K, \cdots, L_{T_k} \leq K|L, \gamma)$ according to Theorem 2, which is a multivariate $t$ distribution with corresponding parameters. Consequently (12) becomes a mixture of $k + 1$ multivariate $t$ distributions, with the last one degenerating to a point mass distribution. Clearly, it requires more computation efforts compared with that in the first scenario, which is the price paid for the unknown change point.
Figure 4: *Modeling of vibration signals* The $x$-axis is the working time of operations, and $y$-axis is the log magnitude of the vibration. The solid line is the fitted model proposed in this paper, and the dashed line is the real degradation signals.

5 Numerical Analysis and Experimental Results

In this section, we first use a real dataset containing the degradation data of rotational bearings as an example to illustrate the overall procedures of the condition monitoring using the proposed framework. Then a simulation study is conducted to quantitatively evaluate the performance of the proposed method.

5.1 Degradation monitoring of rotational bearings

We consider the condition monitoring of the rotational bearings, as discussed in the introduction. The dataset was first studied by Gebraeel et al. (2005), and subsequently used in (e.g. Gebraeel 2006; Gebraeel and Lawley 2008). Twenty-five bearings with complete records of the vibration signals are available. Examples of the collected signals are shown in Figure 1. When the bearing vibrates with magnitude exceeding 0.03, it fails as suggested in (Gebraeel et al. 2005).

5.1.1 Prior estimation

The prior distributions of model parameters are estimated from existing datasets. Figure 4 illustrates a fitted degradation model of one bearing. It shows that the estimated model can fit the degradation data adequately, which illustrates the suitability of our model in this context. Using the MLE of the model parameters ($\hat{\theta}_i, \hat{\gamma}_i$) and the form of the prior distributions specified in Section
Table 1: Estimated hyper-parameters of the prior distributions

<table>
<thead>
<tr>
<th>Hyper-parameters</th>
<th>Phase 1 Model</th>
<th>Phase 2 Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi(\beta_m</td>
<td>\sigma_m^2)$</td>
<td>$\mu_m$</td>
</tr>
<tr>
<td></td>
<td>$\Sigma_m$</td>
<td>$[-1.40 \times 10^{-1}, -1.43 \times 10^{-4}]$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$[-1.43 \times 10^{-4}, -9.13 \times 10^{-6}]$</td>
</tr>
<tr>
<td>$\pi(\sigma_m^2)$</td>
<td>$v$</td>
<td>$3.66$</td>
</tr>
<tr>
<td></td>
<td>$s^2$</td>
<td>$7.27 \times 10^{-3}$</td>
</tr>
</tbody>
</table>

3, we can estimate the hyper-parameters of the prior distributions, as summarized in Table 1. Table 1 lists the hyperparameters of $\pi(\beta_m | \sigma_m^2)$ and $\pi(\sigma_m^2)$ in both phases respectively. Not surprisingly, the model of the second phase has larger intercept and slope, which indicates faster degradation. Moreover, the variability of the vibration magnitude is also noticeably larger in the second phase, which is consistent with our observations shown from Figure 1.

5.1.2 RUL prediction

With the estimated prior distributions and updating rules in Section 4, we can predict the vibration magnitude at any future time. As pointed out in Theorem 2, the joint distribution of degradation predictions at a set of future times follows multivariate $t$ distribution. In this example, we choose $\mathcal{T} = \{4, 8, 12, \cdots, 4k, \cdots\}$ as possible failure times. Two scenarios of predictions are considered here as introduced in Section 4.2: predictions made before the change point and predictions made after the change point. First we consider the performance when we predict before the change point, i.e., only using observations from the first phase. Figure 5 summarizes the prediction intervals of the RUL of all 25 bearings when predicted at two time points, corresponding to 75% and 90% of the real change point time of each bearing respectively. Although some of the predictions miss the real failure times, many of them still have acceptable performance considering the information we have from the limited degradation observations and the large variability of the change points.

In contrast, predictions made after change points are expected to be more accurate. Since similar work has been done on the prediction of RUL using the second phase degradation (e.g. Gebraeel et al. 2005), we also compare the performance of our method with theirs (named GLLR for short). The essential difference is that GLLR assumed the predicted vibration magnitude at
The $\circ$ denotes the 5%, 50%, and 95% quantile of the RUL distributions; $\ast$ is the actual failure time.

\[
P(R_T \leq T_k - \tau | L) = 1 - P(LT_k \leq K | L).
\]

Figure 6 illustrates an example of computed RUL distribution predicted at two different times, corresponding to 75% and 90% of the actual failure time. Although the RUL is discretely distributed, we use the density-like plot instead of the probability mass distribution for easier comparison. The real RUL at these two time points are 169 and 67 respectively. Figure 6 reveals that our predicted distribution of the RUL is tighter and more accurate compared with that by GLLR. And the prediction variance is smaller when the updating time is closer to the actual failure time, when more observations are available. For a comprehensive evaluation of our methods, we construct the confidence intervals of the failure time of all the bearings. Similar to the methods in the literature, we report 5%, 50%, and 95% quantiles of the RUL distribution. Figure 7 shows the prediction intervals as well as actual failure times of all bearings. From the figure, we can find that in general when the prediction is made at the time closer to the actual failure time, the prediction interval is smaller, and tends to be more accurate. Compared with the results provided in the literature, our prediction intervals are tighter and has better coverage probability. Noticeably, in contrast to Figure 5, the prediction intervals in Figure 7 are much more accurate and tighter. This reflects the benefits provided by additional degradation observations and more accurate estimation of the
Figure 6: Example of predicted distribution of the RUL. The x-axis is the RUL, and y-axis is the corresponding probability. The solid line corresponds to the proposed method, and the dashed line is the GLLR method. The vertical dotted line indicates the actual RUL at the time of predictions.

5.2 Performance evaluation using simulation

In this part, we use simulation to further evaluate the performance of our condition monitoring method. We use the length and coverage probability of the prediction interval of RUL as evaluation criteria, which are important and commonly used in the literature. Given the same confidence level, the shorter prediction interval indicates more accurate predictions. At the same time, the real coverage probability of the prediction interval should match the designed confidence level. In the simulation, we assume the degradation path follows the two-phase model in (2) exactly. The location of the change point $\gamma$ follows shifted exponential distribution:

$$f(\gamma) = \begin{cases} \exp[-(t - 200)/150]/150, & t \geq 200 \\ 0, & \text{otherwise} \end{cases}$$

The hyper-parameters of the prior distributions of the model parameters are listed in Table 1. Initially, 50 random sample paths are generated from the model to estimate the prior distributions as in Section 3. Subsequently, we use the estimated prior distributions and the updating/predicting procedures in Section 4 to estimate the distribution of the RUL based on newly generated degra-
dation data.

5.2.1 Accuracy of predictions after change point

In this part, we evaluate the performance when we predict the RUL after the change point. In this scenario, the GLLR method is also applicable. As we mentioned before, their method does not consider the correlations among predicted values. We can assess how much improvement we can make by taking this effect into consideration.

In each simulation replication, degradation data are generated according to model (2). The parameters of the model are randomly sampled from the specified (unrevealed) distribution. We consider three prediction times, which correspond to 50%, 75%, and 90% of the time between degradation change point and failure time. Figure 8 compares the prediction intervals (at confidence level 0.1) in seven simulation replications between the proposed method and the GLLR method. It shows that our prediction intervals can cover the actual RUL time more accurately with much smaller interval. Additionally, the intervals become narrower when the predictions are made at a later stage (with more observations).

We also use 1000 replications to compute the coverage probability as well as the length of the prediction intervals. If the prediction intervals are valid, their coverage probability should be consistent with the confidence level, which equals 0.9 in our simulation. Additionally, if the length of
Figure 8: Examples of prediction intervals at different time points. The numbers in the x-axis represent the replication number. The solid lines indicate the prediction interval using proposed method, and the dashed lines correspond to the prediction interval from the literature. The * markers are the actual RUL.
Table 2: Coverage probability of prediction intervals at three time points

<table>
<thead>
<tr>
<th>Method</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>0.821</td>
<td>0.851</td>
<td>0.827</td>
</tr>
<tr>
<td>GLLR</td>
<td>0.312</td>
<td>0.277</td>
<td>0.197</td>
</tr>
</tbody>
</table>

Figure 9: Distribution of the length of prediction intervals. The solid line is the density of the length of the prediction interval using proposed method, and the dashed line is the corresponding density using the GLLR method.

The prediction interval is shorter, the RUL prediction is less uncertain. Table 2 compares the coverage probability of the prediction interval obtained at three prediction points. The proposed method performs much better than the GLLR method in terms of coverage probability. It slightly misses the designed confidence level, which might be caused by the approximation using the empirical Bayes method and the countable set of failure times. The lengths of their prediction intervals are compared in Figure 9. The densities of the interval length are estimated from 1000 simulation replications. Clearly the our method has shorter prediction intervals than that of GLLR in general. The prediction interval also becomes shorter and more informative as more observations are available. We would like to emphasize that Figure 9 is not the density of the RUL, which shall decrease much faster as observation time increases, as illustrated in Figure 6. All the evidence from simulation results support that the prediction accuracy of the RUL can be improved significantly by considering the variability of error terms and the correlations among predicted values.
5.2.2 Accuracy of predictions before change point

In this part, we evaluate the performance of predictions that are made before the change points. This is one of the advantageous features not offered by existing methods. We use the same simulation settings as in Section 5.2.1, and set the prediction time at 75% and 90% of their corresponding change point times. It is interesting to see how accurate the prediction performs without accurate information regarding the change point of the degradation. Figure 10 illustrates the prediction interval in different simulation replications. The prediction intervals are much wider in this case compared with that in Figure 8. Not surprisingly, different replications provide similar prediction intervals. This is intuitive because the degradation observations in the first phase have very small variabilities. And the RUL mainly depends on the degradations in the second phase, whose model parameters follow the same prior distribution in this circumstance. Nevertheless, the prediction interval still provides a rough range of the failure times, and has consistent confidence level as designed.

6 Conclusion

In this paper, we propose a two-phase degradation model for condition monitoring of rotational bearings. We use the Bayesian framework to efficiently integrate the degradation information from
historical data with in-situ observations of each new unit in operation to provide accurate prediction of degradation magnitudes at future time points. In addition, we explicitly consider the correlation among the multiple predictions to improve the accuracy of RUL distribution. The advantages of our method have been demonstrated using extensive numerical studies from both real dataset and simulation experiments.

Nevertheless, there are also some open issues worthy further investigation. First, in current model, we assume the location of change point does not have significant effects on other degradation parameters. However, as revealed from the real data, when the change point occurs at a later stage, the degradation of the second phase tends to be faster. Therefore, we may get a better estimate of the RUL by considering this correlated effects between the change point and the degradation speed. Second, we may consider flexible number of degradation phases in the model to be applicable to other types of degradation data. This feature might be extremely important in the applications where physical knowledge of the degradation processes is limited. Third, extending the approach to other general nonlinear degradation models also worths further research to enrich its applications.

**Acknowledgment**

We would like to thank the reviewers and the editor for their constructive comments. Nan Chen is partially supported by the AcRF funding R-266-000-057-133, Kwok-Leung Tsui is partially supported by CityU SRG #7002553, RGC CRF #CityU8/CRF/09, Start-up projects #CityU9676001 and #CityU9380048.
Appendix

A.1 Proof of Theorem 1

Proof From the model (2) and (7), we can find that $Y_1$, $Y_2$ are completely determined by $\hat{\gamma}$, and thus $P(L|\theta, \hat{\gamma})$ can be written as $P(Y_1|\beta_1, \sigma_1, \hat{\gamma}) \cdot P(Y_2|\beta_2, \sigma_2, \hat{\gamma})$. Therefore, from (10), we have

$$P(\theta|L, \hat{\gamma}) \propto P(Y_1|\beta_1, \sigma_1, \hat{\gamma}) \cdot P(Y_2|\beta_2, \sigma_2, \hat{\gamma}) \cdot \pi(\beta_1, \sigma_1) \cdot \pi(\beta_2, \sigma_2)$$

$$\propto P(\beta_1, \sigma_1|Y_1, \hat{\gamma}) \cdot P(\beta_2, \sigma_2|Y_2, \hat{\gamma})$$

$$\propto P(\beta_1, \sigma_1|L, \hat{\gamma}) \cdot P(\beta_2, \sigma_2|L, \hat{\gamma}), \quad (A.1)$$

which indicates the independence between the two posterior distributions given a constant $\hat{\gamma}$. Additionally, their symmetric structure indicates the posterior distributions of the parameters in each phase have the same form. In particular, for each $m = 1, 2$

$$P(\beta_m, \sigma_m|Y_m) \propto \sigma_m^{-n_m} \exp \left[ -\frac{\|Y_m - X_m\beta_m\|^2}{2\sigma_m^2} \right] \cdot \sigma_m^{-v_m-2} \exp \left( -\frac{v_m s_m^2}{2\sigma_m^2} \right)$$

$$\cdot \sigma_m^{-2} \exp \left[ -\frac{1}{2\sigma_m^2} (\beta_m - \mu_m)^T \Sigma^{-1}_m (\beta_m - \mu_m) \right]$$

$$\propto \sigma_m^{-2} \exp \left[ -\frac{\beta_m^T (X_m^T X_m + \Sigma^{-1}_m) \beta_m - 2(Y_m^T X_m + \mu_m^T \Sigma^{-1}_m) \beta_m}{2\sigma_m^2} \right]$$

$$\cdot \sigma_m^{-(n_m + v_m + 2)} \cdot \exp \left[ -\frac{v_m s_m^2 + Y_m^T Y_m + \mu_m^T \Sigma^{-1}_m \mu_m}{2\sigma_m^2} \right]$$

$$\propto \sigma_m^{-2} \exp \left[ -\frac{(\beta_m - \bar{\mu}_m)^T (X_m^T X_m + \Sigma^{-1}_m) (\beta_m - \bar{\mu}_m)}{2\sigma_m^2} \right]$$

$$\cdot \sigma_m^{-(n_m + v_m + 2)} \cdot \exp \left[ -\frac{(v_m + n_m) s_m^2}{2\sigma_m^2} \right]$$

$$= N(\bar{\mu}_m, \sigma_m^2 (X_m^T X_m + \Sigma^{-1}_m)^{-1}) \cdot SI \chi^2(v_m + n_m, s_m^2), \quad (A.2)$$

where the last $\propto$ follows from the matrix operation by constructing the quadratic form of $\beta_m$. 

A.2 Proof of Theorem 2

Proof Based on the degradation model and the definition of RUL, we have

$$P(R_{T_k} \leq T_k - \tau | L) = 1 - P(R_{T_k} > T_k - \tau | L) = 1 - P(L_{T_1} \leq K, L_{T_2} \leq K, \ldots, L_{T_k} \leq K | L),$$
where $L_{T_j}$ is the degradation level at time $T_j$. What remains is to find the joint distribution of $\tilde{L} = [L_{T_1}, L_{T_2}, \cdots, L_{T_k}]^T$. Since the degradation has entered the second phase, we have $\tilde{L} = \tilde{X}\beta_2 + \sigma_2 \epsilon$ according to (2). Therefore, for fixed $\sigma_2$, $\tilde{L}$ follows multivariate normal distribution with mean $\mu$ and covariance $\sigma_2^2 \Sigma$, where

$$\mu \equiv E(\tilde{L}|\sigma_2) = X\beta_2 + \sigma_2 E\epsilon = \tilde{X}\mu_2,$$

$$\sigma_2^2 \Sigma \equiv var(\tilde{L}|\sigma_2) = X var\beta_2 X^T + \sigma_2^2 var\epsilon = \sigma_2^2 [X(X^T X_2 + \Sigma_2^{-1})^{-1}X^T + I].$$

By integrating out the parameter $\sigma_2$, we can obtain the conditional distribution of $\tilde{L}$ given $L$, i.e.,

$$P(\tilde{L}|L) = \int_0^\infty \int P(\tilde{L}|\beta_2, \sigma_2^2) \cdot P(\beta_2|\sigma_2^2, L) \cdot P(\sigma_2^2|L)d\beta_2d\sigma_2^2 = \int_0^\infty P(\tilde{L}|\sigma_2^2) \cdot P(\sigma_2^2|L)d\sigma_2^2$$

$$= \int_0^\infty \exp \left[ -\frac{(L-\mu)^T \Sigma^{-1}(L-\mu)}{2\sigma^2} \right] \cdot \frac{\tilde{v}_2 \Sigma_2^2/2}{\Gamma(\tilde{v}_2/2)} \exp \left[ \frac{-\tilde{v}_2 \Sigma_2^2/2}{2\sigma^2} \right] d\sigma^2$$

$$= \frac{(\tilde{v}_2 \Sigma_2^2/2)^{\tilde{v}_2/2}}{(2\pi)^{k/2}(\det \Sigma)^{1/2}} \cdot \frac{\Gamma(\tilde{v}_2/2)}{\Gamma(\tilde{v}_2+k/2)} \cdot \frac{\exp \left[ -\frac{(L-\mu)^T \Sigma^{-1}(L-\mu) + \tilde{v}_2 \Sigma_2^2/2}{2\sigma^2} \right]}{\sigma^{\tilde{v}_2+k/2}}$$

$$= MT(\mu, \Sigma_2^2, \tilde{v}_2).$$

(A.3)

Consequently, the RUL distribution is $P(R_k - T_k \leq \tau | L) = 1 - P(\tilde{L} \leq K | L)$, which completes the proof.

References


Biography

Nan Chen is an Assistant Professor in the Department of Industrial and Systems Engineering at National University of Singapore. He obtained his B.S. degree in Automation from Tsinghua University, and M.S. degree in Computer Science, M.S. degree in Statistics, and Ph.D. degree in Industrial Engineering from University of Wisconsin-Madison. His research interests include statistical modeling and surveillance of service systems, simulation modeling design, condition monitoring and degradation modeling. He is a member of INFORMS and IIE.

Kwok-Leung Tsui is chair professor and head of the Systems Engineering and Engineering Management at the City University of Hong Kong. He was a professor in the School of Industrial and Systems Engineering at the Georgia Institute of Technology prior to that. He currently serves as Quality and Reliability Departmental Editor of IIE Transactions. He had also held various chair positions in the INFORMS Quality, Statistics, and Reliability Section and Data Mining Section. He is a Fellow of ASA, and member of ASQ and INFORMS. His current research interests include data mining, systems and health informatics, quality control, and computer experiment modeling.