IIE Transactions (2013) **45**, 939–952

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/0740817X.2012.706376

Condition monitoring and remaining useful life prediction using degradation signals: revisited

NAN CHEN[1,*] and KWOK LEUNG TSUI[2]

[1]*Department of Industrial and Systems Engineering, National University of Singapore, 21 Lower Kent Ridge Road, 119077, Singapore*
E-mail: isecn@nus.edu.sg
[2]*Department of Systems Engineering and Engineering Management, City University of Hong Kong, Tat Chee Avenue, Hong Kong*

Received June 2011 and accepted June 2012

Condition monitoring is an important prognostic tool to determine the current operation status of a system/device and to estimate the distribution of the remaining useful life. This article proposes a two-phase model to characterize the degradation process of rotational bearings. A Bayesian framework is used to integrate historical data with up-to-date *in situ* observations of new working units to improve the degradation modeling and prediction. A new approach is developed to compute the distribution of the remaining useful life based on the degradation signals, which is more accurate compared with methods reported in the literature. Finally, extensive numerical results demonstrate that the proposed framework is effective and efficient.

Keywords: Condition monitoring, degradation, remaining useful life, Bayesian

1. Introduction

Modern engineering systems are overwhelmingly complex because of increasing requirements on their functionalities and qualities. These systems often have a high standard of 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. Unfortunately, traditional ways to predict system failures often fail in practice (Pecht, 2008) because knowledge about the failure mechanism 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. *In situ* sensing data are often collected and analyzed during system operations to characterize and predict a system's health condition (Nelson, 1990), to prepare necessary preventive maintenance (Wang, 2007; Chen *et al.*, 2011; Wang *et al.*, 2012), or even to plan burn-in for future products (Ye, Shen, and Xie, 2012; Ye, Xie, Tang, and Shen, 2012). Built on these real-time sensing data, condition monitoring methods use statistical models to emulate the physical degradation processes. They have been shown to be more flexible and widely applicable compared with physical models that rely on thorough knowledge of the failure mechanisms. Methods in condition monitoring can be classified into two categories depending on whether or not the health state of the systems are directly observable. In the first category where the health state is not directly observable, different models (Kumar and Klefsj, 1994; Baruah and Chinnam, 2005) 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, evolution of the health condition is often directly characterized using random coefficient models, Brownian motion, gamma processes, etc. (Doksum and Hyland, 1992; Lu and Meeker, 1993; Park and Padgett, 2005; Si *et al.*, 2012; Ye, Xie, Tang, and Chen, 2012). Si *et al.* (2011) provided an excellent review of both categories of methods to predict the RUL using condition monitoring data.

In this article, we investigate the condition monitoring and RUL prediction of rotational bearings. The bearing degradation is manifested by the magnitude of vibration during rotation. After the bearing is installed, it will experience an initial stable stage, in which the vibration is slight. However, its degradation becomes more and more severe after an unknown change point, and the vibration magnitude increases dramatically and has large variability. The bearing is considered to have failed once its

*Corresponding author

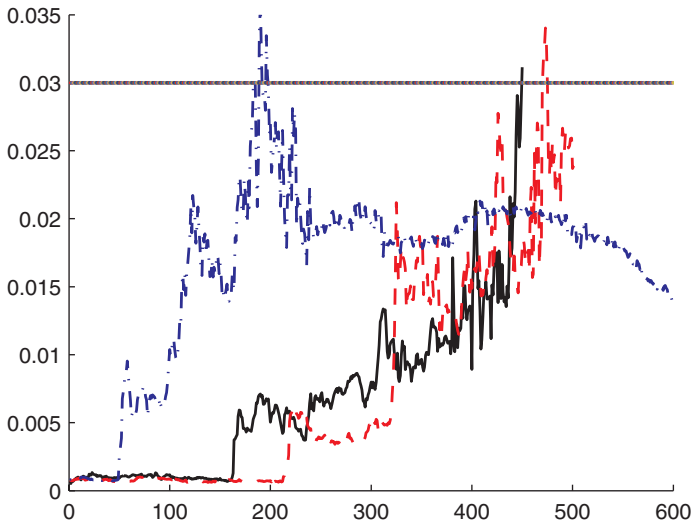

Fig. 1. Vibration signals of three bearings. The x-axis is the working time of operations and the y-axis is the magnitude of the vibration (average of signal magnitudes at seven different harmonic frequencies as used in Gebraeel *et al.* (2005)). The horizontal dotted line represents the failure threshold (color figure provided online).

degradation exceeds a predetermined threshold. However, different bearings often have different change-point locations and large variability in the increment rates of vibration magnitude, despite the similar shapes of their degradation paths, as illustrated in Fig. 1. Gebraeel *et al.* (2005), among other notable works, proposed a Bayesian framework to model this type of degradation signal. They used an exponential model with random coefficients to characterize the evolution of the degradation signal after the change point:

$$L_j \equiv \log[S(t_j) - \delta] = a + b \times t_j + \epsilon_j, \qquad (1)$$

where $S(t_j)$ is the vibration magnitude observed at time t_j, δ is a known constant, a and b are model coefficients (parameters), and ϵ_j is the corresponding error term. In the Bayesian framework they developed, a and b are random variables that follow certain prior distributions, and ϵ_j follows an independent and identically distributed (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 $P(a, b|L_1, L_2, \ldots, L_n)$ 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) and Chakraborty *et al.* (2009).

There are several aspects of current approaches that can be improved. First, existing works often start condition monitoring after the change point of the degradation as if the exact location of the change point is known. However, the change points usually vary between units significantly, as clearly illustrated in Fig. 1. By considering the distribution of the change point, we may be able to predict the RUL right after its installment. Second, the variance of ϵ_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, again demonstrated in Fig. 1. Consequently, appropriate modeling of the variance heterogeneity is also expected to improve the prediction accuracy. Third and most important, 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 is not considered appropriately, the distribution of RUL might be inaccurate. In this article, we improve these aspects while modeling the degradation of bearings. 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 the 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 non-linear, 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 technical details on how to update the model parameters and predict the RUL of a new working unit using the Bayesian method. Section 5 demonstrates the effectiveness of the proposed method using numerical simulations and examples from a real dataset. Section 6 concludes the article with discussions on future research.

2. Two-phase degradation model

In this article, 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 (Loader, 1991; Lin, 2008; Yuan and Kuo, 2010). In our approach, we adopt a similar piecewise log-linear model as that used by Bae and Kvam (2006), but we use the Bayesian approach to update the model parameters and make predictions.

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

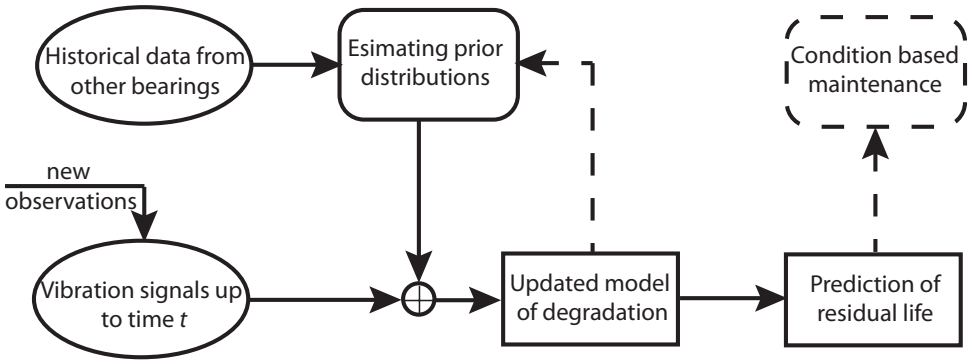

Fig. 2. Flowchart of model updating and RUL prediction in condition monitoring. The dashed line and box indicate other useful tasks that are not included in this article.

degradation model as

$$L_{ij} = \ln[S(t_{ij}) - \delta] = \begin{cases} a_{i1} + b_{i1}t_{ij} + \sigma_{i1}\epsilon_{ij}, & t_{ij} \le \gamma_i \\ a_{i2} + b_{i2} \times (t_{ij} - \gamma_i) + \sigma_{i2}\epsilon_{ij}, & t_{ij} > \gamma_i \end{cases}, \quad (2)$$

where ϵ_{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; γ_i is the random variable denoting the change point of unit i. By setting $\gamma_i = 0$, and σ_{i2} the same for all i, our model reduces to the simpler model considered in Gebraeel *et al.* (2005) and Gebraeel (2006). From Fig. 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 $\boldsymbol{\beta}_{i1} = [a_{i1}, b_{i1}]^{\mathrm{T}}$, $\boldsymbol{\beta}_{i2} = [a_{i2}, b_{i2}]^{\mathrm{T}}$, $\boldsymbol{\theta}_i = [\boldsymbol{\beta}_{i1}^{\mathrm{T}}, \sigma_{i1}, \boldsymbol{\beta}_{i2}^{\mathrm{T}}, \sigma_{i2}]^{\mathrm{T}}$ 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 information on the degradation of failed units provides information on the possible values of the model parameters $\boldsymbol{\theta}_i, \gamma_i$ of the same type of units. The *in situ* sensing observations make the model parameters more and more accurate toward the true value of the particular working unit. In more detail, we assume the model parameters $\boldsymbol{\theta}_i, \gamma_i$ are random variables following a common prior distribution with density $\pi(\boldsymbol{\theta}_i, \gamma_i)$, which can be learned from historical data. When sensing observations from a new working unit i are available up to current time τ, we can update the posterior distribution of the parameters $\boldsymbol{\theta}_i, \gamma_i$ using the Bayes formula (Gelman *et al.*, 2004):

$$P(\boldsymbol{\theta}_i, \gamma_i | \mathbf{L}_i(\tau)) \propto P(\mathbf{L}_i(\tau) | \boldsymbol{\theta}_i, \gamma_i) \times \pi(\boldsymbol{\theta}_i, \gamma_i), \quad (3)$$

where $\mathbf{L}_i(\tau) \equiv [L_{i1}, L_{i2}, \ldots, L_{ij}], \forall t_{ij} \le \tau$ are the observations of the degradation magnitude, and $P(\mathbf{L}_i(\tau) | \boldsymbol{\theta}_i, \gamma_i)$ is the likelihood function based on model (2). Using the sensing information $\mathbf{L}_i(\tau)$ allows the model with updated parameters to better fit the degradation of unit i. 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:

$$P(L_{ik} | \mathbf{L}_i(\tau)) = \int P(L_{ik} | \boldsymbol{\theta}_i, \gamma_i, \mathbf{L}_i(\tau)) \times P(\boldsymbol{\theta}_i, \gamma_i | \mathbf{L}_i(\tau)) \mathrm{d}\boldsymbol{\theta}_i \mathrm{d}\gamma_i. \quad (4)$$

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

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

3. Specification of priors

Prior information is a unique and important component in the Bayesian framework. It offers the opportunity to effectively integrate domain knowledge or past experience with newly observed data. 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 can be estimated from historical data. In this article, we take the objective approach by estimating the priors from historical degradation data of failed units, in a similar manner to Gebraeel (2006).

Assuming the prior distribution can be expressed as a parametric model $\pi(\boldsymbol{\theta}, \gamma | \boldsymbol{\zeta})$ with hyper-parameters $\boldsymbol{\zeta}$, the empirical Bayes method can be adopted to estimate these

hyper-parameters by maximizing the marginal likelihood:

$$\hat{\zeta} = \arg\max_{\zeta} \prod_{i=1}^{I} \int P(\mathbf{L}_i|\theta_i, \gamma_i) \times \pi(\theta_i, \gamma_i|\zeta) \mathrm{d}\theta_i \mathrm{d}\gamma_i,$$

where $\mathbf{L}_i = [L_{i1}, L_{i2}, \ldots, L_{in_i}]^{\mathrm{T}}$ is the historical degradation of unit i, n_i is the number of observations of degradation, and θ_i, γ_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 article, we propose an alternative approach to estimate the prior distribution. Instead of maximizing the marginal distribution of the historical data, we consider the Maximum Likelihood Estimates (MLEs) of their model parameters $\hat{\theta}_i, \hat{\gamma}_i$ of each unit for $i = 1, 2, \ldots, I$, as samples from the prior distributions. Subsequently, we can estimate the prior distributions using distribution selection and fitting techniques such as histograms or χ^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 γ_i. By definition, the log-likelihood function given observations $\mathbf{L}_i$ of unit i can be written as

$$\begin{aligned} l(\theta_i, \gamma_i|\mathbf{L}_i) &= \sum_{j=1}^{n_i} \left[-\frac{1}{2} \ln\left(2\pi\sigma_{i1}^2\right) - \frac{(L_{ij} - a_{i1} - b_{i1}t_{ij})^2}{2\sigma_{i1}^2} \right] \\ &\quad \times \mathcal{I}(t_{ij} \le \gamma_i) \\ &\quad + \sum_{j=1}^{n_i} \left[-\frac{1}{2} \ln\left(2\pi\sigma_{i2}^2\right) - \frac{(L_{ij} - a_{i2} - b_{i2}t_{ij} + b_{i2}\gamma_i)^2}{2\sigma_{i2}^2} \right] \\ &\quad \times \mathcal{I}(t_{ij} > \gamma_i), \end{aligned} \tag{5}$$

where $\mathcal{I}(\cdot)$ is the indicator function, which equals one when the condition is true and zero otherwise. It is difficult to directly maximize Equation (5) due to the discontinuities. Instead, we can consider the likelihood conditioned on γ_i; i.e., $l(\theta_i|\mathbf{L}_i, \gamma_i)$. With fixed γ_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 obtain the closed-form solution of $\boldsymbol{\beta}_{im}$ and σ_{im} for $m = 1, 2$ that maximizes the $l(\theta_i|\mathbf{L}_i, \gamma_i)$ as

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{im} &= \left(\mathbf{X}_{im}^{\mathrm{T}}\mathbf{X}_{im}\right)^{-1}\mathbf{X}_{im}^{\mathrm{T}}\mathbf{Y}_{im}, \\ \hat{\sigma}_{im}^2 &= \frac{1}{n_{im}}(\mathbf{Y}_{im} - \mathbf{X}_{im}\hat{\boldsymbol{\beta}}_{im})^{\mathrm{T}}(\mathbf{Y}_{im} - \mathbf{X}_{im}\hat{\boldsymbol{\beta}}_{im}), \ m = 1, 2, \end{aligned} \tag{6}$$

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

$$\begin{aligned} &\mathbf{Y}_{i1} = [L_{i1}, L_{i2}, \ldots, L_{i\lfloor\gamma_i\rfloor}]^{\mathrm{T}}, \quad \mathbf{Y}_{i2} = [L_{i\lceil\gamma_i\rceil}, \ldots, L_{in_i}]^{\mathrm{T}} \\ &\mathbf{X}_{i1} = \begin{bmatrix} 1, & 1, & \ldots, & 1 \\ t_{i1}, & t_{i2}, & \ldots, & t_{i\lfloor\gamma_i\rfloor} \end{bmatrix}^{\mathrm{T}}, \\ &\mathbf{X}_{i2} = \begin{bmatrix} 1, & \ldots, & 1, & 1 \\ t_{i\lceil\gamma_i\rceil} - \gamma_i, & \ldots, & t_{i(n_i-1)} - \gamma_i, & t_{in_i} - \gamma_i \end{bmatrix}^{\mathrm{T}} \end{aligned} \tag{7}$$

where $\lfloor\gamma_i\rfloor$ ($\lceil\gamma_i\rceil$) are the largest (smallest) integer such that $t_{i\lfloor\gamma_i\rfloor} \le \gamma_i$ ($t_{i\lceil\gamma_i\rceil} > \gamma_i$). It is worth mentioning here that the quantities $\hat{\boldsymbol{\beta}}_{im}, \hat{\sigma}_{im}, n_{im}$ are functions of γ_i. Here we do not explicitly express this dependence to avoid tedious notation. The conditional maximum log-likelihood, which is defined as $M(\gamma_i) = \max_{\theta_i} l(\theta_i|\mathbf{L}_i, \gamma_i)$, can therefore be expressed as

$$\begin{aligned} M(\gamma_i) = -\sum_{m=1}^{2} \Bigg[&\frac{n_{im}}{2} + \frac{n_{im}}{2} \\ &\times \ln \frac{2\pi \left\| \mathbf{Y}_{im} - \mathbf{X}_{im}\left(\mathbf{X}_{im}^{\mathrm{T}}\mathbf{X}_{im}\right)^{-1}\mathbf{X}_{im}^{\mathrm{T}}\mathbf{Y}_{im} \right\|^2}{n_{im}} \Bigg], \end{aligned} \tag{8}$$

where $\|\cdot\|^2$ denotes the $\mathcal{L}_2$-norm of a vector. It can be proved that $\hat{\gamma}_i = \arg\max M(\gamma_i)$ together with $\hat{\theta}_i = \arg\max l(\theta_i|\mathbf{L}_i, \hat{\gamma}_i)$ jointly maximize the likelihood function in Equation (5). From the derivation, we can find that $\mathbf{X}_{im}, \mathbf{Y}_{im}$ only change when γ_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, $\hat{\gamma}_i$ exists and is finite. We would like to point out that there also exist other methods to determine the location proposed by γ_i, such as those of Lai (1995) and Siegmund and Venkatraman (1995). However, to limit this scope of this article, we do not further discuss alternative methods.

We can use the maximum likelihood estimates $(\hat{\theta}_i, \hat{\gamma}_i)$ for all the units $i = 1, 2, \ldots, I$ that failed in the past to estimate the prior distribution. In our model, the prior distribution is multi-dimensional and requires a 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:

$$\pi(\theta_i, \gamma_i) = \pi(\gamma_i)\pi(\boldsymbol{\beta}_{i1}, \sigma_{i1}|\gamma_i)\pi(\boldsymbol{\beta}_{i2}, \sigma_{i2}|\gamma_i). \tag{9}$$

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 the degradation model is highly desired. Therefore, in this article, we also prefer using conjugate priors for efficient model updating. Based on our model, we can specify the priors as

$$\begin{aligned} \pi(\boldsymbol{\beta}_{im}, \sigma_{im}|\gamma_i) &= \pi\left(\boldsymbol{\beta}_{im}|\sigma_{im}^2, \gamma_i\right)\pi\left(\sigma_{im}^2|\gamma_i\right) \\ &= N\left(\boldsymbol{\mu}_m, \sigma_m^2\boldsymbol{\Sigma}_m\right) \times SI\chi^2\left(v_m, s_m^2\right), \quad m = 1, 2, \end{aligned}$$

where $N(\boldsymbol{\mu}_m, \sigma_m^2\boldsymbol{\Sigma}_m)$ is a normal distribution with density function:

$$f\left(\boldsymbol{\beta}; \boldsymbol{\mu}_m, \sigma_m^2\boldsymbol{\Sigma}_m\right) = \frac{1}{(2\pi)^{\kappa/2}\sigma_m^{\kappa/2}(\det \boldsymbol{\Sigma}_m)^{1/2}} \times \exp\left[-\frac{(\boldsymbol{\beta}-\boldsymbol{\mu}_m)^{\mathrm{T}}\boldsymbol{\Sigma}_m^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_m)}{2\sigma_m^2}\right],$$

where κ is the dimension of $\boldsymbol{\beta}$. $SI\chi^2(v_m, s_m^2)$ is the scaled inverse χ^2 distribution with density:

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

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

Compared with the distribution $\pi(\boldsymbol{\beta}_{im}, \sigma_{im}|\gamma_i)$, $\pi(\gamma_i)$ 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 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 that different units work independently, their updating and prediction procedures are the same and can be done independently. Therefore, we omit the unit index i hereafter in this section. In addition, we consider the updating and prediction at a single time τ. However, 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 τ, we can update the degradation model of the unit by computing the posterior distribution of its parameter $\boldsymbol{\theta}, \gamma$. 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 $0 \le t_1 < t_2 < \cdots < t_n < \tau$ as $\mathbf{L} = [L_1, L_2, \ldots, L_n]^{\mathrm{T}}$. Using the Bayes formula we have

$$P(\boldsymbol{\theta}, \gamma|\mathbf{L}) \propto P(\mathbf{L}|\boldsymbol{\theta}, \gamma) \times \pi(\boldsymbol{\theta}, \gamma).$$

However, computing this posterior distribution is time-consuming, which is not suitable for 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 $\tilde{\gamma}$ of $P(\gamma|\mathbf{L})$. In the second step, we consider $\tilde{\gamma}$ as the real change point and perform the updating of $\boldsymbol{\theta}$ using $P(\boldsymbol{\theta}|\mathbf{L}, \tilde{\gamma})$. In practice, as the number of observations increases, $P(\gamma|\mathbf{L})$ concentrates more and more around $\tilde{\gamma}$. Therefore, this approximation becomes more accurate in subsequent updates. Figure 3 illustrates the posterior distribution $P(\gamma|\mathbf{L})$ at two different updating times. In the left-hand part, the model is updated when the degradation is still in the first phase. The posterior mode is equal to the last observation time t_n, which indicates that no change point is suspected in the previous observations. In the right-hand part, the model is updated when the degradation has already entered the second phase. The posterior mode is able to find the real change point.

After $\tilde{\gamma}$ has been identified, we can further compute the posterior distribution of $\boldsymbol{\theta}$. Instead of integrating out γ given the unit's degradation observations $\mathbf{L}$, we use a point mass distribution at $\tilde{\gamma}$ to approximate the conditional distribution $P(\gamma|\mathbf{L})$; that is,

$$\begin{aligned} P(\boldsymbol{\theta}|\mathbf{L}) &= \int P(\boldsymbol{\theta}, \gamma|\mathbf{L})\mathrm{d}\gamma \\ &= \int \frac{P(\mathbf{L}|\boldsymbol{\theta}, \gamma) \times \pi(\boldsymbol{\theta}|\gamma) \times \pi(\gamma)}{P(\mathbf{L})}\mathrm{d}\gamma \\ &= \int \frac{P(\mathbf{L}|\boldsymbol{\theta}, \gamma) \times \pi(\boldsymbol{\theta}|\gamma)}{P(\mathbf{L}|\gamma)} \times P(\gamma|\mathbf{L})\mathrm{d}\gamma \\ &\simeq \frac{P(\mathbf{L}|\boldsymbol{\theta}, \tilde{\gamma}) \times \pi(\boldsymbol{\theta}|\tilde{\gamma})}{P(\mathbf{L}|\tilde{\gamma})}. \end{aligned} \tag{10}$$

The last approximation in Equation (10) is valid because $P(\gamma|\mathbf{L})$ concentrates around $\tilde{\gamma}$ with a reasonable sample size, as shown in Fig. 3. Using this empirical Bayes approximation, the posterior distribution of $\boldsymbol{\theta}$ can be computed using Theorem 1 (the proof is included in Appendix A1).

Theorem 1. *Given $\tilde{\gamma}$, the posterior distributions $P(\boldsymbol{\beta}_1, \sigma_1|\mathbf{L}, \tilde{\gamma})$ and $P(\boldsymbol{\beta}_2, \sigma_2|\mathbf{L}, \tilde{\gamma})$ are independent. And for $m = 1, 2$, $P(\sigma_m^2|\mathbf{L}, \tilde{\gamma})$ follows scaled inverse χ^2 distribution with parameter $\tilde{v}_m = v_m + n_m$ and*

$$\tilde{s}_m^2 = \left[v_m s_m^2 + \mathbf{Y}_m^{\mathrm{T}}\mathbf{Y}_m + \boldsymbol{\mu}_m^{\mathrm{T}}\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu}_m - \tilde{\boldsymbol{\mu}}_m^{\mathrm{T}}(\mathbf{X}_m^{\mathrm{T}}\mathbf{X}_m + \boldsymbol{\Sigma}_m^{-1})\tilde{\boldsymbol{\mu}}_m\right]/(v_m + n_m).$$

And $P(\boldsymbol{\beta}_m|\sigma_m^2, \mathbf{L})$ follows normal distribution with mean $\tilde{\boldsymbol{\mu}}_m = (\mathbf{X}_m^{\mathrm{T}}\mathbf{X}_m + \boldsymbol{\Sigma}_m^{-1})^{-1}(\mathbf{X}_m^{\mathrm{T}}\mathbf{Y}_m + \boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu}_m)$ and variance $\sigma_m^2(\mathbf{X}_m^{\mathrm{T}}\mathbf{X}_m + \boldsymbol{\Sigma}_m^{-1})^{-1}$. The matrices $\mathbf{X}_m$, $\mathbf{Y}_m$ are determined through Equation (7) with γ_i substituted by $\tilde{\gamma}$.

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 that when $\tilde{\gamma}$ is equal to the

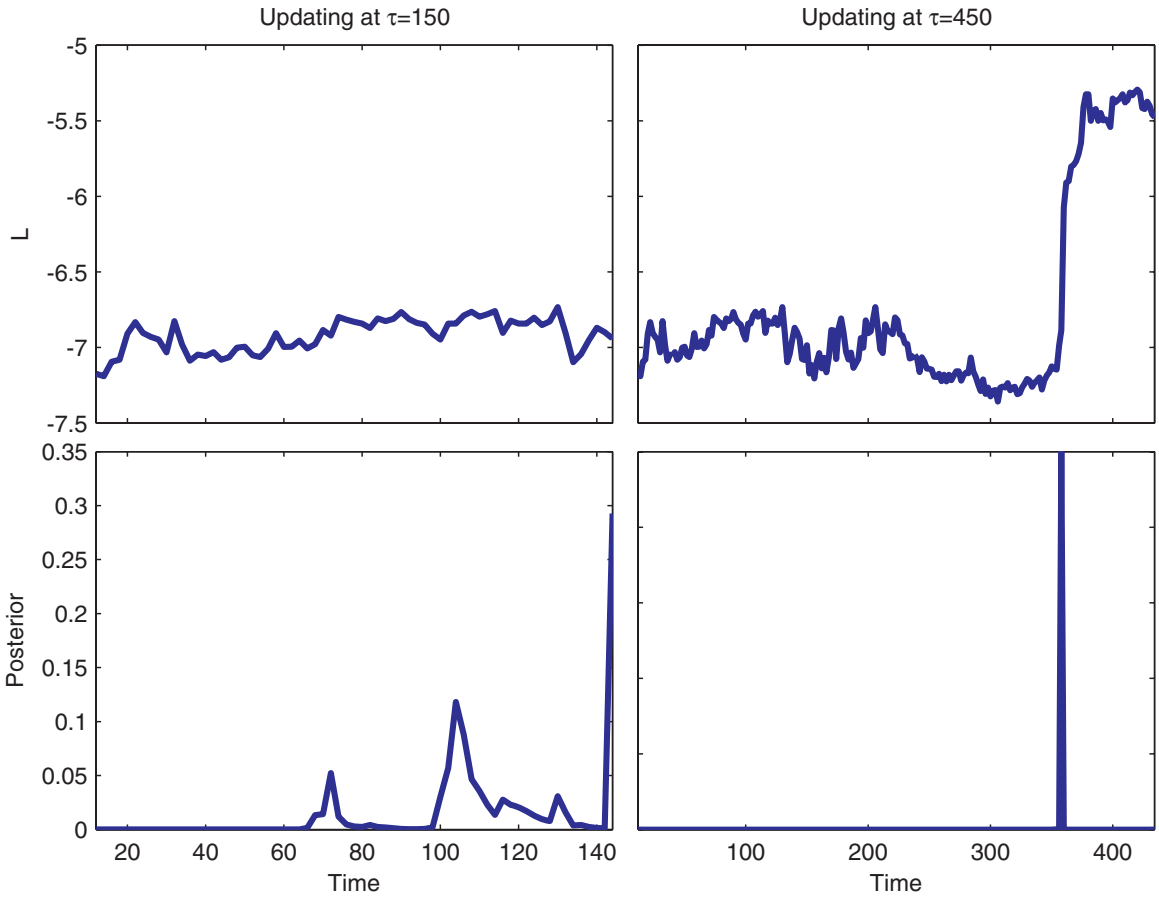

Fig. 3. Posterior probability of the location of the change point at different times. The upper panels are the degradation observations up to two time points, and the bottom panels are the corresponding posterior probability $P(\gamma|\mathbf{L})$ (color figure provided online).

last observation time, i.e., $\tilde{\gamma} = t_n$, we have $\mathbf{Y}_1 = \mathbf{L}, \mathbf{Y}_2 = 0$. Therefore, it is not necessary to update the parameters in the second phase since no observations from there are available. However, this does not imply that 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, $P(\gamma|\mathbf{L})$ cannot be approximated by a degenerated point mass distribution at $\tilde{\gamma}$. It might be approximated by the conditional distribution $\pi(\gamma|\gamma > \tilde{\gamma})$ instead. Although this difference will not influence the updating rule of $\boldsymbol{\theta}$, it will change the prediction of RUL, as discussed in Section 4.2.

4.2. *RUL prediction*

After the degradation model of the working unit is 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 ϵ 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 $\tilde{\gamma}$. These two scenarios have different prediction procedures. First, we consider the case when $\tilde{\gamma} < t_n$, which means the degradation has entered in the second phase at time t_n. If the empirical Bayes approximation is valid—in other words, $\tilde{\gamma}$ is close to the real change point—the prediction only depends on $\gamma, \boldsymbol{\beta}_2, \sigma_2$ 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 the 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 Appendix A2). It is worth noting 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 a practical point of view even if the failure time is continuously distributed.

Theorem 2. *Given the collection of possible failure times after current time* τ, $\mathfrak{T}_\tau = \{T_k | k = 1, 2, \ldots, T_k > \tau \geq \tilde{\gamma}\}$, *the RUL distribution* $P(R_\tau \leq T_k - \tau | \mathbf{L}) = 1 - MT_k(\mathbf{K})$ *where* R_τ *denotes the RUL at time* τ; $\mathbf{K}$ *is k-dimensional vector with each entry K denoting the failure threshold;* $MT_k(\cdot)$ *is the cumulative distribution function of a k-dimensional multivariate t distribution with degree of freedom* $\tilde{v}_2$, *mean* $\bar{\mathbf{X}}\tilde{\boldsymbol{\mu}}_2$ *and squared scale matrix* $\tilde{s}_2^2[\mathbf{I} + \bar{\mathbf{X}}(\mathbf{X}_2^{\mathrm{T}}\mathbf{X}_2 + \boldsymbol{\Sigma}_2^{-1})^{-1}\bar{\mathbf{X}}^{\mathrm{T}}]$. $\bar{\mathbf{X}}$ *is the matrix defined as*

$$\bar{\mathbf{X}}^{\mathrm{T}} = \begin{bmatrix} 1, & 1, \ldots, & 1 \\ T_1 - \tilde{\gamma}, & T_2 - \tilde{\gamma}, \ldots, & T_k - \tilde{\gamma} \end{bmatrix}_{2\times k}.$$

Theorem 2 provides a way to compute the distribution of RUL at any given time point τ. 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., $\tilde{\gamma} = t_n$, 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 $P(\gamma|\mathbf{L})$. 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:

$$\begin{aligned} P(L_{T_k}|\mathbf{L}) = &\int_{T_k}^{\infty} P(L_{T_k}|\boldsymbol{\beta}_1, \sigma_1) \times P(\boldsymbol{\beta}_1, \sigma_1|\mathbf{L}, \gamma) \\ &\times \pi(\gamma|\gamma > \tilde{\gamma})\mathrm{d}\gamma \\ &+ \int_{\tilde{\gamma}}^{T_k} P(L_{T_k}|\boldsymbol{\beta}_2, \sigma_2) \times P(\boldsymbol{\beta}_2, \sigma_2|\mathbf{L}, \gamma) \\ &\times \pi(\gamma|\gamma > \tilde{\gamma})\mathrm{d}\gamma. \end{aligned} \tag{11}$$

Since the observations $\mathbf{L}$ are only available up to $t_n = \tilde{\gamma}$, the posterior distributions of the model parameters $\boldsymbol{\theta}$ will be the same given $\mathbf{L}$ regardless the true change point γ. Consequently, according to the similar derivation in the proof of Theorem 2, Equation (11) becomes a mixture of a t-distribution with weight $\omega_1 = \int_{T_k}^{\infty} \pi(\gamma|\gamma > \tilde{\gamma})\mathrm{d}\gamma$ 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 a similar reasoning still applies.

In practice, as demonstrated in Fig. 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 almost certainly larger than γ. Therefore, the RUL over the set of possible failure times after τ, $\mathfrak{T}_\tau = \{T_k | k = 1, 2, \ldots, T_k > \tau\}$, can be computed as

$$\begin{aligned} &P(R_\tau > T_k - \tau|\mathbf{L}) \\ &\quad = P(L_{T_1} \leq K, L_{T_2} \leq K, \ldots, L_{T_k} \leq K|\mathbf{L}) \\ &\quad \simeq \sum_{s=1}^{k} \int_{T_{s-1}}^{T_s} P(L_{T_s} \leq K, L_{T_{s+1}} \leq K, \ldots, L_{T_k} \leq K|\mathbf{L}, \gamma) \\ &\qquad \times \pi(\gamma|\gamma > \tilde{\gamma})\mathrm{d}\gamma + \int_{T_k}^{\infty} \pi(\gamma|\gamma > \tilde{\gamma})\mathrm{d}\gamma, \end{aligned} \tag{12}$$

where we define $T_0 = \tilde{\gamma}$ for notational simplicity and use L_{T_s} to denote the degradation at time T_s. We can compute $\mathrm{P}(L_{T_s} \leq K, L_{T_{s+1}} \leq K, \ldots, L_{T_k} \leq K|\mathbf{L}, \gamma)$ using Theorem 2, which is a multivariate t distribution with corresponding parameters. Consequently, Equation (12) becomes a mixture of $k + 1$ multivariate t distributions, with the last one degenerating to a point mass distribution. Clearly, it requires more computational effort compared with that in the first scenario, which is the price paid for the unknown change point.

5. Numerical analysis and experimental results

In this section, we first use a real dataset containing degradation data on 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 rotational bearings, as discussed in the Introduction. The dataset was first studied by Gebraeel *et al.* (2005) and subsequently used in Gebraeel (2006) and Gebraeel and Lawley (2008). Twenty-five bearings with complete records of their vibration signals are available. Examples of the collected signals are shown in Fig. 1. When the bearing vibrates with a magnitude that exceeds 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 $(\hat{\boldsymbol{\theta}}_i, \hat{\gamma}_i)$ and the form of the prior distributions specified in Section 3, we can estimate the hyper-parameters of the prior distributions, as summarized in Table 1. Table 1 lists the hyperparameters of $\pi(\boldsymbol{\beta}_m|\sigma_m^2)$ and $\pi(\sigma_m^2)$ in both phases, respectively. Not surprisingly, the model of the second phase has a larger intercept and slope, which indicates faster degradation.

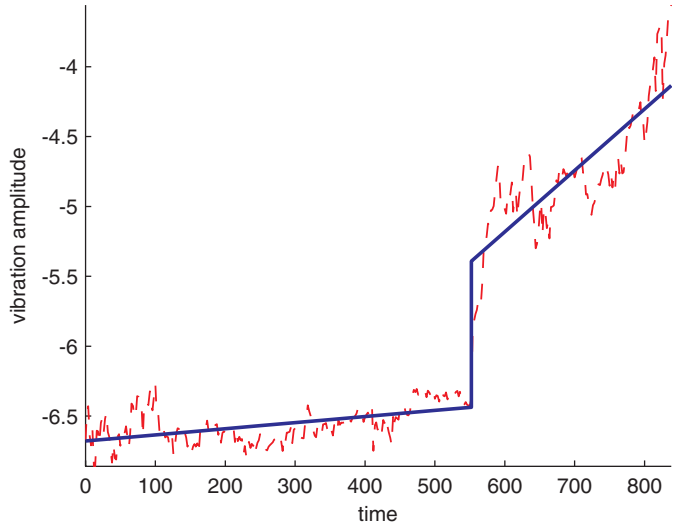

Fig. 4. Modeling of vibration signals. The x-axis is the working time of operations, and the y-axis is the log magnitude of the vibration. The solid line is the fitted model proposed in this article, and the dashed line is the real degradation signals (color figure provided online).

Moreover, the variability of the vibration magnitude is also noticeably larger in the second phase, which is consistent with our observations shown in Fig. 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 a multivariate t distribution. In this example, we choose $\mathfrak{T} = \{4, 8, 12, \ldots, 4k, \ldots\}$ 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 an 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 different time points are monotone. Hence,

$$P(R_\tau \le T_k - \tau|\mathbf{L}) = 1 - P(L_{T_k} \le K|\mathbf{L}).$$

Figure 6 illustrates an example of computed RUL distributions 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 RULs 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. Also, 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 constructed 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 a 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 have a better coverage probability. Noticeably, in contrast with Fig. 5, the prediction intervals in Fig. 7 are much more accurate and tighter. This reflects the benefits provided by additional degradation observations and more accurate estimation of the change point.

Table 1. Estimated hyper-parameters of the prior distributions

Hyper-parameters		*Phase 1 model*	*Phase 2 model*
$\pi(\boldsymbol{\beta}_m\|\sigma_m^2)$	$\boldsymbol{\mu}_m$	$[-7.11, 1.48 \times 10^{-5}]$	$[-5.19, 3.85 \times 10^{-3}]$
	$\boldsymbol{\Sigma}_m$	$\begin{bmatrix} 1.40 \times 10^{-1} & -1.43 \times 10^{-4} \\ -1.43 \times 10^{-4} & 9.13 \times 10^{-6} \end{bmatrix}$	$\begin{bmatrix} 2.06 \times 10^{-3} & -5.47 \times 10^{-4} \\ -5.47 \times 10^{-4} & 3.79 \times 10^{-6} \end{bmatrix}$
$\pi(\sigma_m^2)$	v	3.66	6.48
	s^2	7.27×10^{-3}	5.46×10^{-2}

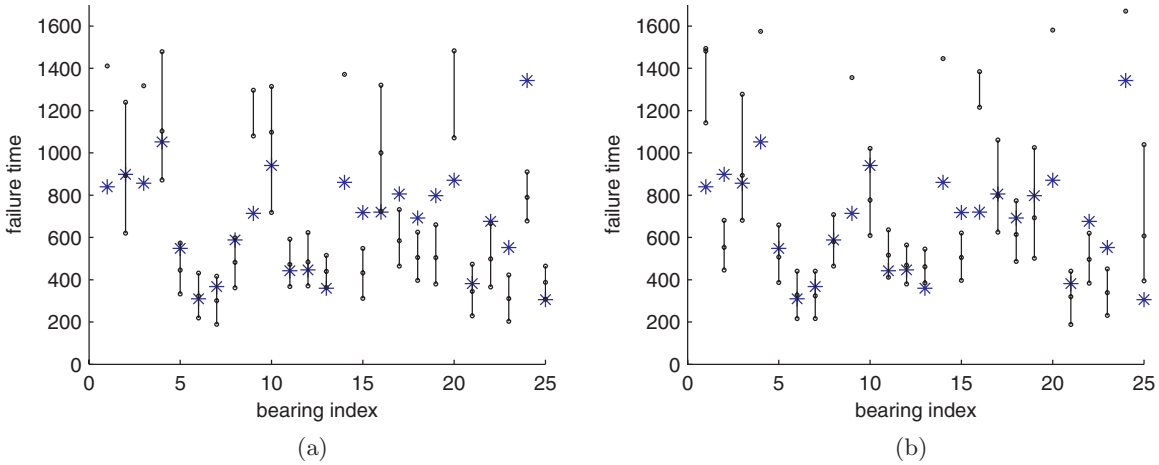

Fig. 5. Prediction interval of RUL obtained before change points (a) at 75% of the change-point time and (b) at 90% of the change-point time. The ○ denotes the 5, 50, and 95% quantile of the RUL distributions; ∗ is the actual failure time (color figure provided online).

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, a shorter prediction interval indicates a more accurate prediction. At the same time, the real coverage probability of the prediction interval should match the designed confidence level. In the simulation, we assumed that the degradation path exactly follows the two-phase model in Equation (2). The location of the change point γ follows a 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 were generated from the model to estimate the prior distributions as in Section 3. Subsequently, we used the estimated prior distributions and the updating/predicting procedures in Section 4 to estimate the distribution of the RUL based on newly generated degradation data.

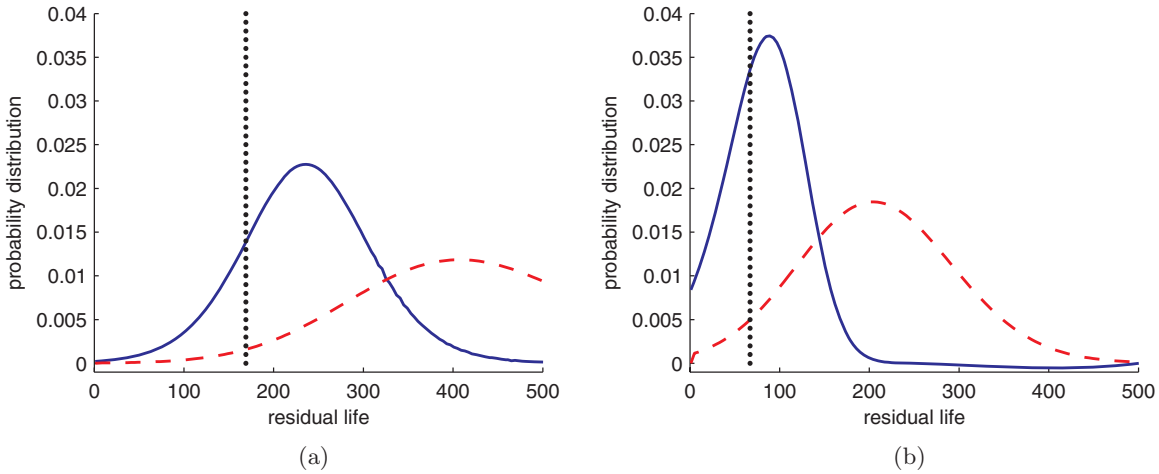

Fig. 6. Examples of predicted distribution of the RUL (a) at 75% of the failure time and (b) at 90% of the failure time. The x-axis is the RUL, and the y-axis is the corresponding probability. The solid line corresponds to the proposed method, and the dashed line is the GLLR method. The vertical dotted line indicates the actual RUL at the time of the prediction (color figure provided online).

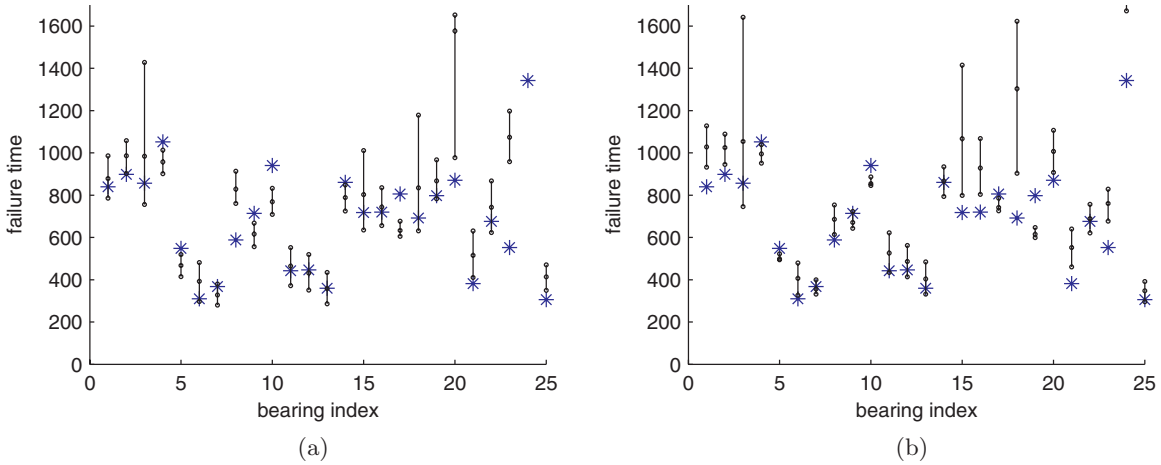

Fig. 7. Prediction intervals of failure times made after change points (a) at 75% of failure time and (b) at 90% of failure time. The ○ denotes the 5, 50, and 95% quantile of the RUL distributions; ∗ is the actual failure time (color figure provided online).

5.2.1. *Accuracy of predictions after the 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 were generated according to model (2). The parameters of the model were randomly sampled from the specified (unrevealed) distribution. We considered 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 a confidence level of 0.9) 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 a much smaller interval. Additionally, the intervals become narrower when the predictions are made at a later stage (with more observations).

We also used 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 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 Fig. 9. The densities of the interval length are estimated from 1000 simulation replications. Clearly, our method generally has shorter prediction intervals than that of GLLR. The prediction interval also becomes shorter and more informative as more observations are available. We would like to emphasize that Fig. 9 is not the density of the RUL, which will decrease much faster as observation time increases, as illustrated in Fig. 6. All of 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 used 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 accurately 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 Fig. 8. Not surprisingly, different replications provide similar prediction intervals. This is intuitive because the degradation observations in the first phase have very small variabilities. Also, the RUL mainly depends on the degradations in the second phase, whose model parameters follows the same prior distribution in this circumstance. Nevertheless, the prediction

Table 2. Coverage probability of prediction intervals at three time points

Method	*50%*	*75%*	*90%*
Proposed	0.821	0.851	0.827
GLLR	0.312	0.277	0.197

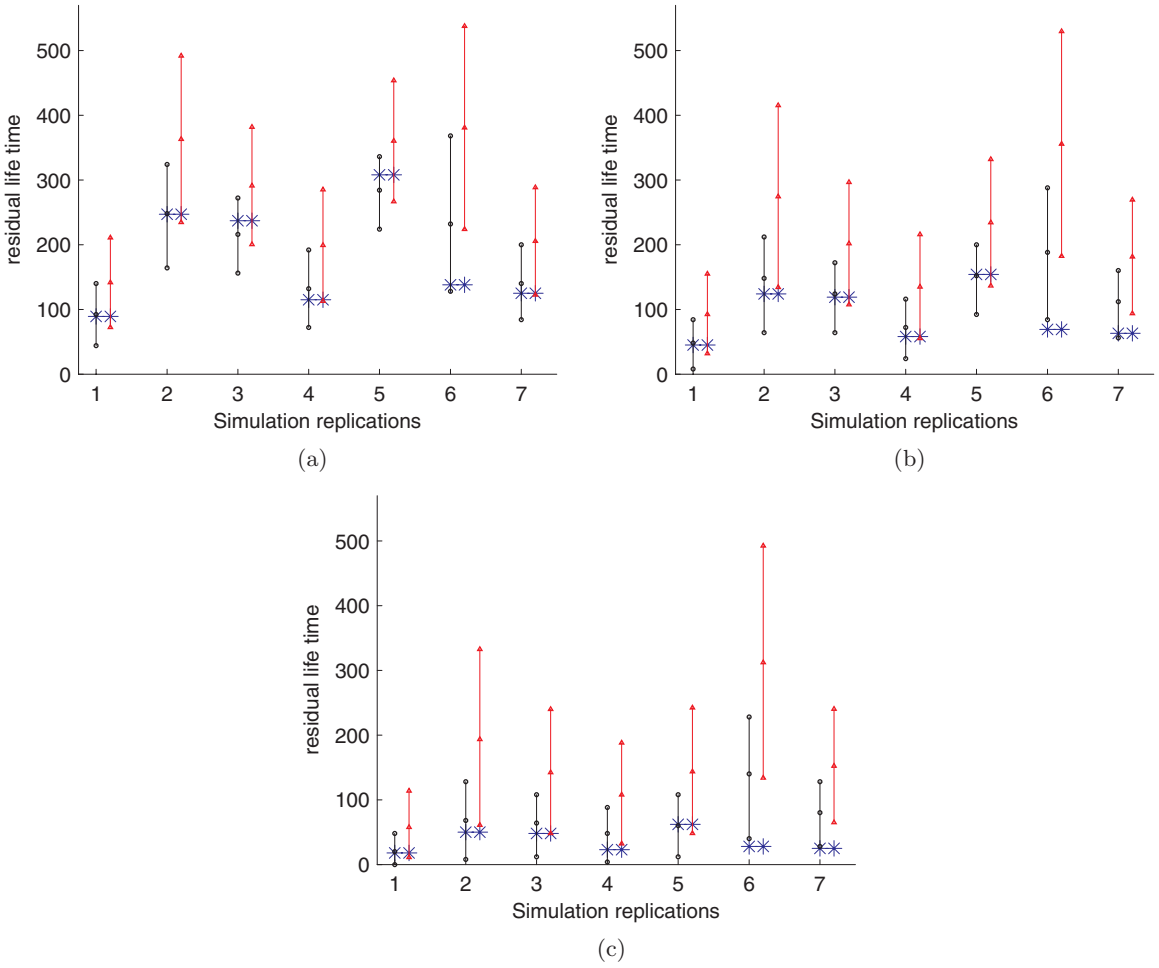

Fig. 8. Examples of prediction intervals at different time points: (a) at 50% failure time; (b) at 75% failure time; and (c) at 90% failure time. The numbers on the *x*-axis represent the replication number. The solid lines indicate the prediction interval using the proposed method, and the dashed lines correspond to the prediction interval from the literature. The * markers are the actual RUL (color figure provided online).

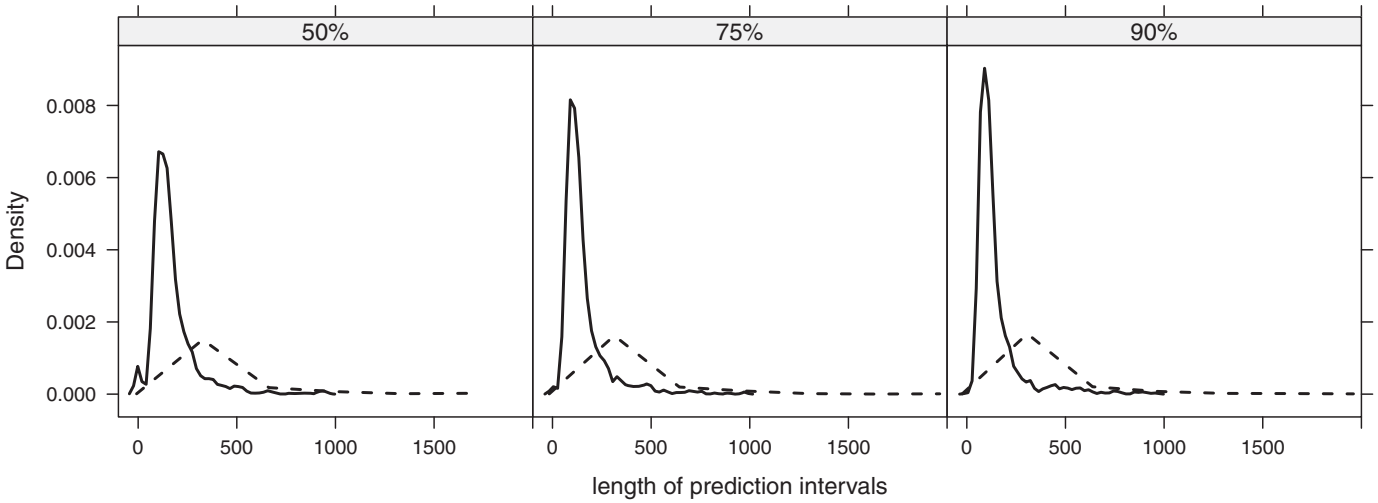

Fig. 9. Distribution of the length of prediction intervals. The solid line is the density of the length of the prediction interval using the proposed method, and the dashed line is the corresponding density using the GLLR method.

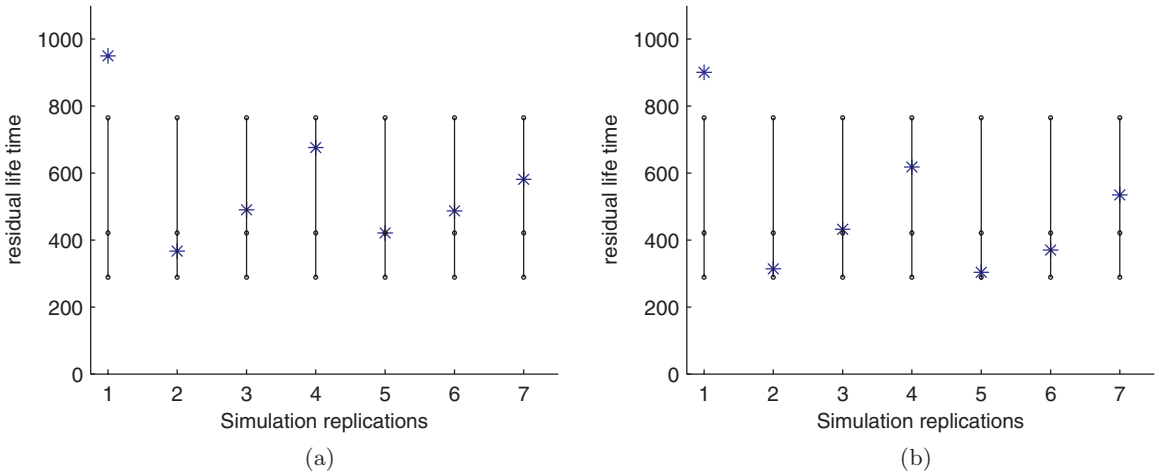

Fig. 10. Examples of prediction intervals at different time points before change points (a) at 75% of change-point time and (b) at 90% of change-point time. The number on the x-axis represents the replication number. The solid lines indicate the prediction interval, and the * markers are the actual RUL (color figure provided online).

interval still provides a rough range of the failure times and has a consistent confidence level as designed.

6. Conclusions

In this article, 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 considered 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 of further investigation. First, in the current model, we assume that the location of the change point does not have a significant effect 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 effect between the change point and the degradation speed. Second, we may consider a flexible number of degradation phases in the model to be applicable to other types of degradation data. This feature might be extremely important in applications where physical knowledge of the degradation processes is limited. Third, extending the approach to other general non-linear degradation models is also worthy of further research to enrich its applications.

Acknowledgements

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, and Kwok-Leung Tsui is partially supported by CityU SRG 7002553, RGC CRF CityU8/CRF/09, and Start-up projects CityU9676001 and CityU9380048.

References

Bae, S.J. and Kvam, P.H. (2006) A change-point analysis for modeling incomplete burn-in for light displays. *IIE Transactions*, **38**, 489–498.

Baruah, P. and Chinnam, R.B. (2005) HMMs for diagnostics and prognostics in machining processes. *International Journal of Production Research*, **43**, 1275–1293.

Carlin, B.P. and Louis, T.A. (2000) *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman & Hall/CRC, New York, NY.

Chakraborty, S., Gebraeel, N., Lawley, M. and Wan, H. (2009) Residual-life estimation for components with non-symmetric priors. *IIE Transactions*, **41**, 372–387.

Chen, N., Chen, Y., Li, Z., Zhou, S. and Sievenpiper, C. (2011) Optimal variability sensitive condition-based maintenance with a Cox PH model. *International Journal of Production Research*, **49**, 2083–2100.

De Boor, C. (2001) *A Practical Guide to Splines*, Springer Verlag, New York, NY.

Doksum, K.A. and Hyland, A. (1992) Models for variable-stress accelerated life testing experiments based on wiener processes and the inverse Gaussian distribution. *Technometrics*, **34**, 74–82.

Gebraeel, N. (2006) Sensory-updated residual life distributions for components with exponential degradation patterns. *IEEE Transactions on Automation Science and Engineering*, **3**, 382–393.

Gebraeel, N., Lawley, M., Li, R. and Ryan, J. (2005) Residual-life distributions from component degradation signals: a Bayesian approach. *IIE Transactions*, **37**, 543–557.

Gebraeel, N.Z. and Lawley, M.A. (2008) A neural network degradation model for computing and updating residual life distributions. *IEEE Transactions on Automation Science and Engineering*, **5**, 154–163.
Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004) *Bayesian Data Analysis*, second edition, Chapman & Hall/CRC, New York, NY.
Kumar, D. and Klefsj, B. (1994) Proportional hazards model: a review. *Reliability Engineering & System Safety*, **44**, 177–188.
Lai, T.L. (1995) Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 613–658.
Lin, J. (2008) A two-stage failure model for Bayesian change point analysis. *IEEE Transactions on Reliability*, **57**, 388–393.
Loader, C.R. (1991) Inference for a hazard rate change point. *Biometrika*, **78**, 749–757.
Lu, C.J. and Meeker, W.Q. (1993) Using degradation measures to estimate a time-to-failure distribution. *Technometrics*, 35, 161–174.
Nelson, W.B. (1990) *Accelerated Testing: Statistical Models, Test Plans, and Data Analysis*, John Wiley & Sons, New York, NY.
Ng, T.S. (2008) An application of the EM algorithm to degradation modeling. *IEEE Transactions on Reliability*, **57**, 2–13.
Park, C. and Padgett, W. (2005) Accelerated degradation models for failure based on geometric brownian motion and gamma processes. *Lifetime Data Analysis*, **11**, 511–527.
Pecht, M.G. (2008) *Prognostics and Health Management of Electronics*, John Wiley & Sons, Hoboken, NJ.
Seber, G.A.F. and Lee, A.J. (2003) *Linear Regression Analysis*, John Wiley & Sons, Hoboken, NJ.
Si, X.-S., Wang, W., Hu, C.-H. and Zhou, D.-H. (2011) Remaining useful life estimation—a review on the statistical data driven approaches. *European Journal of Operational Research*, **213**, 1–14.
Si, X., Wang, W., Hu, C., Zhou, D. and Pecht, M. (2012) Remaining useful life estimation based on a nonlinear diffusion degradation process. *IEEE Transactions on Reliability*, **61**, 50–67.
Siegmund, D. and Venkatraman, E.S. (1995) Using the generalized likelihood ratio statistic for sequential detection of a change-point. *The Annals of Statistics*, **23**, 255–271.
Wang, W. (2007) A two-stage prognosis model in condition based maintenance. *European Journal of Operational Research*, **182**, 1177–1187.
Wang, W., Hussin, B. and Jefferis, T. (2012) A case study of condition based maintenance modelling based upon the oil analysis data of marine diesel engines using stochastic filtering. *International Journal of Production Economics*, **136**, 84–92.
Ye, Z., Shen, Y. and Xie, M. (2012) Degradation-based burn-in with preventive maintenance. *European Journal of Operational Research*, **221**, 360–367.
Ye, Z., Xie, M., Tang, L. and Shen, Y. (2012) Degradation-based burn-in planning under competing risks. *Technometrics*, **54**, 159–168.
Ye, Z.S., Xie, M., Tang, L.C. and Chen, N. (2012) Efficient semiparametric estimation of gamma processes for deteriorating products. *Technometrics*, in press.
Yuan, T. and Kuo, Y. (2010) Bayesian analysis of hazard rate, change point, and cost-optimal burn-in time for electronic devices. *IEEE Transactions on Reliability*, **59**, 132–138.

Appendix

A1. *Proof of Theorem 1*

Proof. From models (2) and (7), we can find that $\mathbf{Y}_1$, $\mathbf{Y}_2$ are completely determined by $\tilde{\gamma}$, and thus $P(\mathbf{L}|\boldsymbol{\theta}, \tilde{\gamma})$ can be written as $P(\mathbf{Y}_1|\boldsymbol{\beta}_1, \sigma_1, \tilde{\gamma}) \times P(\mathbf{Y}_2|\boldsymbol{\beta}_2, \sigma_2, \tilde{\gamma})$. Therefore, from Equation (10), we have

$$\begin{aligned} P(\boldsymbol{\theta}|\mathbf{L}, \tilde{\gamma}) &\propto P(\mathbf{Y}_1|\boldsymbol{\beta}_1, \sigma_1, \tilde{\gamma}) \times P(\mathbf{Y}_2|\boldsymbol{\beta}_2, \sigma_2, \tilde{\gamma}) \\ &\quad \times \pi(\boldsymbol{\beta}_1, \sigma_1) \times \pi(\boldsymbol{\beta}_2, \sigma_2) \\ &\propto P(\boldsymbol{\beta}_1, \sigma_1|\mathbf{Y}_1, \tilde{\gamma}) \times P(\boldsymbol{\beta}_2, \sigma_2|\mathbf{Y}_2, \tilde{\gamma}) \\ &\propto P(\boldsymbol{\beta}_1, \sigma_1|\mathbf{L}, \tilde{\gamma}) \times P(\boldsymbol{\beta}_2, \sigma_2|\mathbf{L}, \tilde{\gamma}), \end{aligned} \tag{A1}$$

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

$$\begin{aligned} &P(\boldsymbol{\beta}_m, \sigma_m|\mathbf{Y}_m) \\ &\propto \sigma_m^{-n_m} \exp\left[-\frac{\|\mathbf{Y}_m - \mathbf{X}_m\boldsymbol{\beta}_m\|^2}{2\sigma_m^2}\right] \\ &\quad \times \sigma_m^{-v_m-2} \exp\left(-\frac{v_m s_m^2}{2\sigma_m^2}\right) \\ &\quad \times \sigma_m^{-2} \exp\left[-\frac{1}{2\sigma_m^2}(\boldsymbol{\beta}_m - \boldsymbol{\mu}_m)^{\mathrm{T}}\boldsymbol{\Sigma}_m^{-1}(\boldsymbol{\beta}_m - \boldsymbol{\mu}_m)\right] \\ &\propto \sigma_m^{-2} \exp \\ &\quad \times \left[-\frac{\boldsymbol{\beta}_m^{\mathrm{T}}\left(\mathbf{X}_m^{\mathrm{T}}\mathbf{X}_m + \boldsymbol{\Sigma}_m^{-1}\right)\boldsymbol{\beta}_m - 2\left(\mathbf{Y}_m^{\mathrm{T}}\mathbf{X}_m + \boldsymbol{\mu}_m^{\mathrm{T}}\boldsymbol{\Sigma}_m^{-1}\right)\boldsymbol{\beta}_m}{2\sigma_m^2}\right] \\ &\quad \times \sigma_m^{-(n_m+v_m+2)} \times \exp\left[-\frac{v_m s_m^2 + \mathbf{Y}_m^{\mathrm{T}}\mathbf{Y}_m + \boldsymbol{\mu}_m\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu}_m}{2\sigma_m^2}\right] \\ &\propto \sigma_m^{-2} \exp\left[-\frac{(\boldsymbol{\beta}_m - \tilde{\boldsymbol{\mu}}_m)^{\mathrm{T}}\left(\mathbf{X}_m^{\mathrm{T}}\mathbf{X}_m + \boldsymbol{\Sigma}_m^{-1}\right)(\boldsymbol{\beta}_m - \tilde{\boldsymbol{\mu}}_m)}{2\sigma_m^2}\right] \\ &\quad \times \sigma_m^{-(n_m+v_m+2)} \times \exp\left[-\frac{(v_m + n_m)\tilde{s}_m^2}{2\sigma_m^2}\right] \\ &= N\left(\tilde{\boldsymbol{\mu}}_m, \sigma_m^2\left(\mathbf{X}_m^{\mathrm{T}}\mathbf{X}_m + \boldsymbol{\Sigma}_m^{-1}\right)^{-1}\right) \times SI\chi^2\left(v_m + n_m, \tilde{s}_m^2\right), \end{aligned} \tag{A2}$$

where the last $\propto$ follows from the matrix operation by constructing the quadratic form of $\boldsymbol{\beta}_m$. ■

A2. *Proof of Theorem 2*

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

$$\begin{aligned} P(R_\tau \le T_k - \tau|\mathbf{L}) &= 1 - P(R_\tau > T_k - \tau|\mathbf{L}) \\ &= 1 - P(L_{T_1} \le K, L_{T_2} \le K, \ldots, \\ &\quad L_{T_k} \le K|\mathbf{L}), \end{aligned}$$

where L_{T_j} is the degradation level at time T_j. What remains is to find the joint distribution of $\bar{\mathbf{L}} = [L_{T_1}, L_{T_2}, \ldots, L_{T_k}]^{\mathrm{T}}$. Since the degradation has entered the second phase, we have $\bar{\mathbf{L}} = \bar{\mathbf{X}}\boldsymbol{\beta}_2 + \sigma_2\epsilon$ according to Equation (2). Therefore, for fixed σ_2, $\bar{\mathbf{L}}$ follows multivariate normal distribution with

mean $\bar{\boldsymbol{\mu}}$ and covariance $\sigma_2^2\bar{\boldsymbol{\Sigma}}$, where:

$$\begin{aligned}\bar{\boldsymbol{\mu}} &\equiv \mathrm{E}(\bar{\mathbf{L}}|\sigma_2) = \bar{\mathbf{X}}\mathrm{E}\boldsymbol{\beta}_2 + \sigma_2\mathrm{E}\epsilon = \bar{\mathbf{X}}\tilde{\mu}_2,\\ \sigma_2^2\bar{\boldsymbol{\Sigma}}_m &\equiv \mathrm{var}(\bar{\mathbf{L}}|\sigma_2) = \bar{\mathbf{X}}\mathrm{var}\boldsymbol{\beta}_2\bar{\mathbf{X}}^{\mathrm{T}} + \sigma_2^2\mathrm{var}\epsilon\\ &= \sigma_2^2\left[\bar{\mathbf{X}}\left(\mathbf{X}_2^{\mathrm{T}}\mathbf{X}_2 + \boldsymbol{\Sigma}_2^{-1}\right)^{-1}\bar{\mathbf{X}}^{\mathrm{T}} + \mathbf{I}\right].\end{aligned}$$

By integrating out the parameter σ_2, we can obtain the conditional distribution of $\bar{\mathbf{L}}$ given $\mathbf{L}$; that is,

$$\begin{aligned}P(\bar{\mathbf{L}}|\mathbf{L}) &= \int_0^\infty\!\!\int P\left(\bar{\mathbf{L}}\middle|\boldsymbol{\beta}_2, \sigma_2^2\right) \times P\left(\boldsymbol{\beta}_2\middle|\sigma_2^2, \mathbf{L}\right)\\ &\quad \times P\left(\sigma_2^2\middle|\mathbf{L}\right)\mathrm{d}\boldsymbol{\beta}_2\mathrm{d}\sigma_2^2\\ &= \int_0^\infty P\left(\bar{\mathbf{L}}\middle|\sigma_2^2\right) \times P\left(\sigma_2^2\middle|\mathbf{L}\right)\mathrm{d}\sigma_2^2\\ &= \int_0^\infty \frac{\exp[-((\bar{\mathbf{L}}-\bar{\boldsymbol{\mu}})^{\mathrm{T}}\bar{\boldsymbol{\Sigma}}^{-1}(\bar{\mathbf{L}}-\bar{\boldsymbol{\mu}}))/2\sigma^2]}{(2\pi)^{k/2}\sigma^{k/2}(\det\bar{\boldsymbol{\Sigma}})^{1/2}}\\ &\quad \times \frac{\left(\tilde{v}_2\tilde{s}_2^2/2\right)^{\tilde{v}_2/2}}{\Gamma(\tilde{v}_2/2)}\frac{\exp\left[-\left(\tilde{v}_2\tilde{s}_2^2\right)/2\sigma^2\right]}{(\sigma^2)^{-(\tilde{v}_2/2+1)}}\mathrm{d}\sigma^2\\ &= \frac{\left(\tilde{v}_2\tilde{s}_2^2/2\right)^{\tilde{v}_2/2}}{(2\pi)^{k/2}(\det\bar{\boldsymbol{\Sigma}})^{1/2}\Gamma(\tilde{v}_2/2)}\int_0^\infty (\sigma^2)^{-(\tilde{v}_2+k+1)}\\ &\quad \times \exp\left[-\frac{(\bar{\mathbf{L}}-\bar{\boldsymbol{\mu}})^{\mathrm{T}}\bar{\boldsymbol{\Sigma}}^{-1}(\bar{\mathbf{L}}-\bar{\boldsymbol{\mu}}) + \tilde{v}_2\tilde{s}_2^2}{2\sigma^2}\right]\mathrm{d}\sigma^2\\ &= \frac{\left(\tilde{v}_2\tilde{s}_2^2/2\right)^{\tilde{v}_2/2}}{(2\pi)^{k/2}(\det\bar{\boldsymbol{\Sigma}})^{1/2}\Gamma(\tilde{v}_2/2)}\\ &\quad \times \frac{\Gamma((\tilde{v}_2+k)/2)\times 2^{(\tilde{v}_2+k)/2}}{\left[\tilde{v}_2\tilde{s}_2^2 + (\bar{\mathbf{L}}-\bar{\boldsymbol{\mu}})^{\mathrm{T}}\bar{\boldsymbol{\Sigma}}^{-1}(\bar{\mathbf{L}}-\bar{\boldsymbol{\mu}})\right]^{(\tilde{v}_2+k)/2}}\\ &= MT\left(\bar{\boldsymbol{\mu}}, \tilde{s}_2^2\bar{\boldsymbol{\Sigma}}, \tilde{v}_2\right). \qquad \text{(A3)}\end{aligned}$$

Consequently, the RUL distribution is $P(R_\tau \le T_k - \tau|\mathbf{L}) = 1 - P(\bar{\mathbf{L}} \le \mathbf{K}|\mathbf{L})$, which completes the proof. ■

Biographies

Nan Chen is an Assistant Professor in the Department of Industrial and Systems Engineering at the 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 the 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, IIE, and IEEE.

Kwok Leung Tsui is Chair Professor and Head of the Systems Engineering and Engineering Management Department at the City University of Hong Kong. Previously, he was a Professor in the School of Industrial and Systems Engineering at the Georgia Institute of Technology. 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 a member of ASQ and INFORMS. His current research interests include data mining, systems and health informatics, quality control, and computer experiment modeling.