

Chapman University

Chapman University Digital Commons

Engineering Faculty Articles and Research

Fowler School of Engineering

4-13-2018

Degradation Modeling and RUL Prediction Using Wiener Process Subject to Multiple Change Points and Unit Heterogeneity

Yuxin Wen

Jianguo Wu

Devashish Das

Tzu-Liang Bill Tseng

Follow this and additional works at: https://digitalcommons.chapman.edu/engineering_articles



Part of the [Other Computer Engineering Commons](#), and the [Other Electrical and Computer Engineering Commons](#)

Degradation Modeling and RUL Prediction Using Wiener Process Subject to Multiple Change Points and Unit Heterogeneity

Comments

NOTICE: this is the author's version of a work that was accepted for publication in *Reliability Engineering & System Safety*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Reliability Engineering & System Safety*, volume 176, in 2018.
<https://doi.org/10.1016/j.ress.2018.04.005>.

The Creative Commons license below applies only to this version of the article.

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Copyright

Elsevier

Degradation Modeling and RUL Prediction using Wiener Process subject to Multiple Change Points and Unit Heterogeneity

Yuxin Wen¹, Jianguo Wu*², Devashish Das³, Tzu-Liang (Bill) Tseng⁴

Abstract

Degradation modeling is critical for health condition monitoring and remaining useful life prediction (RUL). The prognostic accuracy highly depends on the capability of modeling the evolution of degradation signals. In many practical applications, however, the degradation signals show multiple phases, where the conventional degradation models are often inadequate. To better characterize the degradation signals of multiple-phase characteristics, we propose a multiple change-point Wiener process as a degradation model. To take into account the between-unit heterogeneity, a fully Bayesian approach is developed where all model parameters are assumed random. At the offline stage, an empirical two-stage process is proposed for model estimation, and a cross-validation approach is adopted for model selection. At the online stage, an exact recursive model updating algorithm is developed for online individual model estimation, and an effective Monte Carlo simulation approach is proposed for RUL prediction. The effectiveness of the proposed method is demonstrated through thorough simulation studies and real case study.

Keywords: Wiener process; multiple change-point model; degradation modeling; remaining useful life prediction

1 Introduction

The fast development of information and sensing technologies offer great opportunities for real-time health condition monitoring and prediction in various modern systems. The condition monitoring signals, also called degradation signals, are commonly used for system reliability assessment due to their direct relation with underlying physical degradation processes [1, 2].

¹ Y. Wen is with the Department of Electrical and Computer Engineering (ECE), University of Texas at El Paso

² J. Wu is with the Department of Industrial, Manufacturing and Systems Engineering, and ECE Department (affiliated), University of Texas at El Paso (e-mail: jwu2@utep.edu).

³ D. Das is with the Department of Industrial and Management Systems Engineering, University of South Florida

⁴ T. Tseng is with the Department of Industrial, Manufacturing and Systems Engineering, University of Texas at El Paso

Typical degradation signals include vibrational signal of rotating machinery [3], crack propagation signal of metallic structures [4], the luminosity of light emitting diode (LED) lamps [5], and the temperature changes and oil debris of engine lubrication [6]. Degradation modeling offers an efficient approach to characterizing the evolution of degradation signals for prognostics, e.g., predicting the remaining useful life (RUL) for an in-service unit based on the available degradation data [7, 8].

In the past few decades, degradation modeling and RUL prediction have been intensively studied. The existing prognostic models can be roughly classified into physics-based and data-driven models [9-11]. The physics-based models requires a thorough understanding of the underlying physical processes that lead to system failure, which is often difficult or infeasible due to large system complexity or unclear degradation mechanisms [12]. In contrast, the data-driven models, which are developed purely on degradation data, are becoming more and more appealing due to unprecedented data availability. In the existing literature, two types of most popular data-driven models are general path models and stochastic process models [13, 14]. The basic idea of the general path models is to use parametric regression to capture how the degradation signal evolves over time. Due to its simplicity and well-established theories, many types of general path models have been built, such as linear or nonlinear regression models with constant or random coefficients [3, 15, 16]. In these models, however, the inherent degradation path is deterministic once the regression parameters are known. It is often oversimplified and is not capable of capturing the temporal uncertainties that are inherent in the degradation process [17, 18]. Therefore, the general path models are applicable only when the temporal uncertainties caused by unobserved internal or external factors are sufficiently small. The stochastic process models, on the other hand, are particularly effective in dealing with such unexplained randomness. The most popular stochastic process models include Gamma process [19], inverse Gaussian process [20], and Wiener process [21]. Due to nice mathematical properties and physical interpretations, Wiener processes have attracted widespread attention. Comprehensive reviews of Wiener process as degradation models can be found in Si *et al.* [16] and Ye *et al.* [17].

Most Wiener process based degradation models assume a linear mean degradation path or can be linearized by time-scale transformations [22]. The key advantage of using a linear drift is that the distribution of the first hitting time (FHT) can be obtained analytically as an inverse Gaussian distribution [23, 24]. From a practical point of view, however, not all degradation signals can be

well modelled by a single mean degradation path function. Some degradation processes show two or even more distinct phases of different drift rates or volatility during the whole life period. Son *et al.* [25] showed that vehicle battery first degrades slowly and then evolves rapidly after the system has degraded down to a certain level before failure occurs. Another representative example is the bearing vibrational signal, where two distinct phases can be easily observed [26]. This phenomenon has also been found in many other systems, such as plasma display panels [27], high-performance capacitors [28], and the liquid coupling devices (LCD) [29]. In all these cases, single-phase models are often not adequate to capture the degradation path evolution.

Motivated by this practical issue, several change-point based general path models have been proposed in recent years. Gebraeel *et al.* [3] chose to truncate observations in the early stage and then fitted an exponential degradation model for the remaining data. Chen and Tsui [26] developed a two-phase regression model with a random change point for the bearing degradation signals. Bae *et al.* [27] proposed a hierarchical Bayesian change-point regression model with one change point for the plasma display panels degradation signals. Different from [26], Bae's model assumes that the two phases are connected, i.e., the degradation path is continuous at the change point. Later, Wen *et al.* [30] proposed a general multiple-phase regression modeling framework for RUL estimation. In their research, a particle filtering algorithm with stratified sampling and partial Gibbs resample-move strategy is developed for online model updating and residual life prediction. As mentioned earlier, the general path models are incapable of capturing the temporal uncertainty of the degradation process. Therefore, incorporating change points to stochastic process for prognostic improvement is desirable.

Wiener process subject to change points has also been studied by several researchers. Ng [31] proposed a Wiener process model with one change point and developed an expectation-maximization (EM) algorithm for model estimation. Kong *et al.* [32] developed a two-phase Wiener processes model considering an abrupt jump at the change point. An EM algorithm was also developed for model parameters. However, both Ng and Kong's methods are not applicable for online condition monitoring and reliability analysis. Feng [28] proposed a two and three-phase Wiener process degradation model to predict the storage lifetime of high-voltage-pulse capacitors for real time reliability analysis. In their work, the locations of change points are assumed to be deterministic for all units. However, this is often not true in many applications. Indeed, the change-point locations often vary from unit to unit, showing significant heterogeneity. Besides, due to

external or internal differences among unites, the degradation rate and volatility may also be heterogeneous, which is not considered in [28]. Recently, Yan *et al.* [33] proposed a two-phase Wiener process degradation model considering the unit heterogeneity for real-time reliability evaluation. However, in their work, how to sequentially estimate the location of the change-point, which is critical for online RUL prediction, is not given.

To overcome these limitations, this paper aims to develop a multiple change-point Wiener process for degradation modeling and RUL prediction. Compared with the existing work, the major contribution of this paper lies in the following aspects: (1) a fully Bayesian approach is formulated to characterize the between-unit heterogeneity in terms of all model parameters, including the change-point locations, drift rate and diffusion; (2) a simple yet effective two-stage empirical approach is proposed for offline model estimation; (3) an efficient recursive model updating algorithm is developed to get the closed-form of the posterior distributions for all model parameters; and (4) an effective Monte Carlo approach is proposed for RUL prediction.

The remainder of this paper is organized as follows. In Section 2, a Wiener process degradation model with multiple change points is presented. The prior parameters specification and estimation are given in Section 3. Section 4 presents the technical details on how to sequentially update the posterior distributions of the current phase, latest change point, and Wiener process parameters of the current phase, and how to predict the RUL. Section 5 demonstrates the effectiveness and accuracy of the proposed method through comprehensive simulation and real case study. The conclusion and discussion are given in Section 6.

2 Wiener Process Degradation Modelling with Multiple Change Points

The general Wiener process can be represented as [34]

$$X(t) = \beta\Lambda(t) + \sigma B(\Lambda(t)) \quad (1)$$

where β is the drift parameter reflecting the rate of degradation, $\sigma > 0$ is called the volatility parameter or diffusion coefficient, $B(\cdot)$ is the standard Brownian motion that captures the stochastic dynamics of the degradation process, and $\Lambda(\cdot)$ is a monotone increasing function representing a general time scale. When $\Lambda(t) = t$, this formula is simplified to the conventional linear Wiener process. Let $\Delta X(t) = X(t + \Delta t) - X(t)$ denote the degradation increments from

time t to $t + \Delta t$. According to the property of Wiener process, the increments are independent and normally distributed as $\Delta X(t) \sim N[\beta[\Lambda(t + \Delta t) - \Lambda(t)], \sigma^2[\Lambda(t + \Delta t) - \Lambda(t)]]$.

In this paper, a multiple change-point Wiener process degradation model is proposed to characterize the degradation path of condition monitoring signals with multiple phases. Specifically, the change points segment the signal into several consecutive phases, where each phase is modelled as a Wiener process. For simplicity, we use simple linear model in each phase for the mean degradation path, i.e., $\Lambda(t) = t$. Suppose the system is inspected at times t_0, t_1, \dots, t_n with degradation observations $X_0 = X(t_0), X_1 = X(t_1), \dots, X_n = X(t_n)$, and assume there are K change points, with index locations c_1, c_2, \dots, c_K . For notational convenience, we define $c_0 = 0$ and $c_{K+1} = n$. Then $c_0 = 0 < c_1 < c_2 < \dots < c_K < c_{K+1} = n$. Therefore, the observations are partitioned into $K + 1$ consecutive phases. Mathematically, the multiple change-point Wiener process can be expressed piecewisely as

$$X(t_j) = \begin{cases} \beta^{(1)}(t_j - t_0) + X(t_0) + \sigma^{(1)}\mathcal{B}(t_j - t_0), & \text{if } t_0 \leq t_j \leq t_{c_1} \\ \beta^{(2)}(t_j - t_{c_1}) + X(t_{c_1}) + \sigma^{(2)}\mathcal{B}(t_j - t_{c_1}), & \text{if } t_{c_1} < t_j \leq t_{c_2} \\ \dots \\ \beta^{(K+1)}(t_j - t_{c_K}) + X(t_{c_K}) + \sigma^{(K+1)}\mathcal{B}(t_j - t_{c_K}), & \text{if } t_{c_K} < t_j \leq t_n \end{cases} \quad (2)$$

where $\beta^{(k)}$ and $\sigma^{(k)}$, $k = 1, \dots, K + 1$ are the drift parameter and diffusion parameter respectively for the k th phase. It is worth noting that in the above model, the mean degradation path is continuous at all change points, which is conventional in the existing literature. It can be easily extended to a general case by adding an extra intercept term for each phase if necessary. Besides, for the sake of simplicity, the starting time of each Wiener process or each phase is assumed to be exactly on the discrete inspection epochs.

To account for the inherent unit-to-unit heterogeneity, the model parameters, including change-point locations, drift rate and diffusion parameter of each phase, are assumed to be random. For the sake of simplicity, the number of change points K is assumed to be deterministic for all units, which is often sufficient for almost all multi-phase degradation signals in the existing literature. If necessary, however, it can be easily extended to a more general case with a random K , as is the case in Wen *et al.* [30].

Bayesian approach is a natural choice to integrate the current available data with historical data for RUL prediction. Under Bayesian framework, the prognostics involves two stages, namely, the offline stage for prior specification and estimation using historical data, and the online stage for

sequential model updating and RUL prediction when new observations are available. There are several remaining challenging issues to address under the multiple change-point framework. First of all, the number of change points K needs to be selected appropriately, which plays a decisive role on the modeling and prediction accuracy. Secondly, the prior distributions for the random model parameters need to be specified, and the corresponding hyperparameters need to be estimated through the historical data. Thirdly, at the online monitoring stage, the posterior distributions of the phase index, the location of the latest change point occurred, and the Wiener process parameters (drift rate and diffusion parameter) have to be sequentially updated once a new observation is available, which is often very challenging. Denote the parameters that need to be updated at the current time index m as a state vector $\boldsymbol{\theta}_m = (\beta_m, \sigma_m^2, \tau_m, s_m)$, where β_m and σ_m^2 are the drift and diffusion parameters of the current phase, τ_m is the latest change point that has occurred ($\tau_m \leq m - 1$), and s_m is the index of the current phase, e.g., $s_m = 1, 2, \dots, K + 1$. Mathematically, the online model updating is to compute the posterior $p(\boldsymbol{\theta}_m | X_{0:m})$, which is highly nonlinear and thus generally intractable. Lastly, based on the updated posterior distributions, we need to predict the residual life. Due to the potential occurrence of future change points, the RUL prediction is very complex. The overall prognostic framework is illustrated in Figure 1. In the following sections, the technical details regarding the aforementioned challenges will be provided.

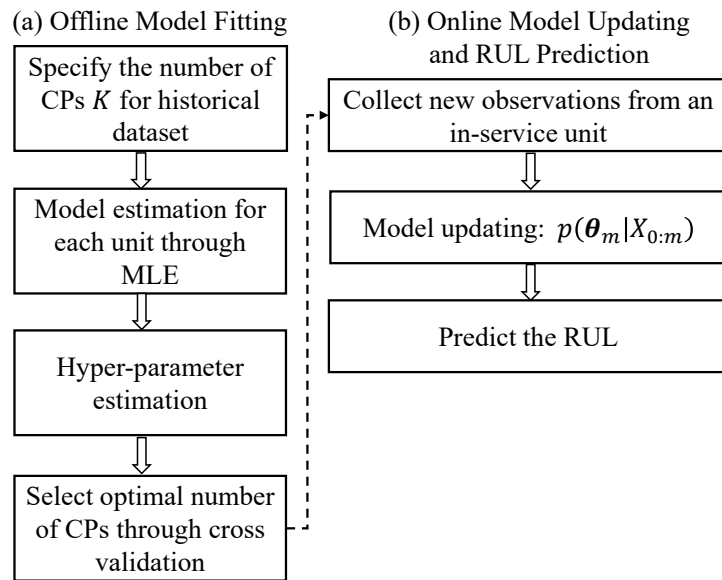


Figure 1. Illustration of the proposed prognostic framework.

3 Offline Prior Specification and Model Estimation

Denote the multiple change-point model as $\mathcal{M} = \left(\{d^{(s)}\}_{s=1}^K, \{\beta^{(s)}\}_{s=1}^{K+1}, \{\sigma^{2(s)}\}_{s=1}^{K+1} \right)$ where K is the number of change points, $d^{(s)}$ is the duration of the s th phase, i.e., $d^{(s)} = t_{c_s} - t_{c_{s-1}}$. In the Bayesian scheme, appropriate priors for \mathcal{M} need to be specified. Prior information, which describes the population-based degradation characteristics, plays a critical role in the posterior inference of a new unit, especially when there are not sufficient observations at the early stage. Instead of directly specifying priors for the change-point locations, we focus on the duration of each phase $d^{(s)}, s = 1, \dots, K$. Note that although there are $K + 1$ phases, only the first K phase durations are needed to identify the change-point locations. For simplicity, we assume that the Wiener process parameters are independent of phase durations, and all phases are independent, except that all phases are piecewise connected. Therefore, the joint prior for phase durations and Wiener process parameters can be formulated as

$$\pi(\mathcal{M}) = \prod_{s=1}^K \pi(d^{(s)}) \prod_{s=1}^{K+1} \pi(\beta^{(s)}, \sigma^{2(s)}) \quad (3)$$

Specifically, we assume that $\pi(d^{(s)})$ follows a normal distribution, i.e., $d^{(s)} \sim N(\mu_d^{(s)}, \sigma_d^{2(s)})$. For $\beta^{(s)}, \sigma^{2(s)}$, a commonly used normal and inverse Gamma (IG) conjugate priors are specified [35],

$$\pi(\beta^{(s)}, \sigma^{2(s)}) = \pi(\sigma^{2(s)})\pi(\beta^{(s)}|\sigma^{2(s)}) = IG(v_0^{(s)}, \gamma_0^{(s)})N(\mu_0^{(s)}, \sigma^{2(s)}\kappa_0^{2(s)}) \quad (4)$$

Luckily, with the above conjugate priors, the joint posterior distribution of θ_m can be exactly calculated sequentially through a recursive modeling updating method, which will be shown later.

Let ψ be the vector of all unknown hyperparameters, i.e., $\psi = \{v_0^{(s)}, \gamma_0^{(s)}, \mu_0^{(s)}, \kappa_0^{2(s)}, \mu_d^{(s)}, \sigma_d^{2(s)}\}$.

In the offline model fitting, all the hyperparameters have to be estimated. Suppose there are I units in the historical dataset. Naturally, the hyperparameters can be obtained from historical dataset by maximizing the following marginal likelihood [26]

$$\hat{\psi} = \arg \max_{\psi} \prod_{i=1}^I \int p(X_i|\mathcal{M}_i)\pi(\mathcal{M}_i|\psi)d\mathcal{M}_i \quad (5)$$

where $\mathbf{X}_i = \{X_{i,1}, \dots, X_{i,n_i}\}$ is the degradation signal for unit i , and n_i is the total number of observations. Unfortunately, the formula is too complex and very difficult to maximize directly. In this paper, we adopt a commonly used empirical two-stage estimation method [25, 26], where the model parameter $\widehat{\mathcal{M}}_i$ for each unit i is estimated at the first stage, and then the hyperparameters are estimated through the maximum likelihood estimation (MLE) by treating $\{\widehat{\mathcal{M}}_i, i = 1, \dots, I\}$ as observations at the second stage.

To take advantage of the independent increment property of $X(t)$, define the increments of observations for unit i as $\delta_{i,1} = X_{i,1} - X_{i,0}$, $\delta_{i,2} = X_{i,2} - X_{i,1}, \dots, \delta_{i,n_i} = X_{i,n_i} - X_{i,n_i-1}$, and time increments as $\lambda_{i,1} = t_{i,1} - t_{i,0}$, $\lambda_{i,2} = t_{i,2} - t_{i,1}, \dots, \lambda_{i,n_i} = t_{i,n_i} - t_{i,n_i-1}$. Then conditioning on \mathcal{M}_i , the increments $\boldsymbol{\delta}_i = (\delta_{i,1}, \delta_{i,2}, \dots, \delta_{i,n_i})'$ follow independent normal distributions given by

$$f(\boldsymbol{\delta}_i | \mathcal{M}_i) = \prod_{s=1}^{K+1} \prod_{j=1}^{n_i^{(s)}} \left(2\pi\sigma_i^{2(s)}\lambda_{i,c_{s-1}+j} \right)^{-\frac{1}{2}} \exp\left(-\frac{(\delta_{i,c_{s-1}+j} - \beta_i^{(s)}\lambda_{i,c_{s-1}+j})^2}{2\sigma_i^{2(s)}\lambda_{i,c_{s-1}+j}} \right) \quad (6)$$

where $n_i^{(s)}$ is the number of observations in the s th phase for unit i . The log-likelihood function can thus be expressed by

$$l(\mathcal{M}_i | \boldsymbol{\delta}_i) = \sum_{s=1}^{K+1} \left[-\frac{n_i^{(s)}}{2} \log(2\pi\sigma_i^{2(s)}) - \frac{1}{2} \sum_{j=1}^{n_i^{(s)}} \log \lambda_{i,c_{s-1}+j} - \frac{1}{2\sigma_i^{2(s)}} \sum_{j=1}^{n_i^{(s)}} \frac{(\delta_{i,c_{s-1}+j} - \beta_i^{(s)}\lambda_{i,c_{s-1}+j})^2}{\lambda_{i,c_{s-1}+j}} \right] \quad (7)$$

Given the change-point locations, the optimal drift and diffusion parameters that maximize Eq. (7) can be obtained as

$$\widehat{\beta}_i^{(s)} = \frac{\sum_{j=1}^{n_i^{(s)}} \delta_{i,c_{s-1}+j}}{\sum_{j=1}^{n_i^{(s)}} \lambda_{i,c_{s-1}+j}}, \quad \widehat{\sigma}_i^{2(s)} = \frac{1}{n_i^{(s)}} \sum_{j=1}^{n_i^{(s)}} \frac{(\delta_{i,c_{s-1}+j} - \widehat{\beta}_i^{(s)}\lambda_{i,c_{s-1}+j})^2}{\lambda_{i,c_{s-1}+j}} \quad (8)$$

Plug in Eq. (8) into (7) we can get a likelihood function with $\{c_{i,s}, s = 1, \dots, K\}$ being the only input variables. Denote $\mathbf{c}_i = \{c_{i,1}, c_{i,2}, \dots, c_{i,K}\}$, $\boldsymbol{\beta}_i = \{\beta_i^{(1)}, \dots, \beta_i^{(K+1)}\}$ and $\boldsymbol{\sigma}_i^2 =$

$\{\sigma_i^{2(1)}, \dots, \sigma_i^{2(K+1)}\}$. The optimal change-point locations can be easily obtained by enumerating all possible values

$$\hat{c}_i = \arg \max_{c_i} l(\mathbf{c}_i, \hat{\boldsymbol{\beta}}_i(\mathbf{c}_i), \hat{\boldsymbol{\sigma}}_i^2(\mathbf{c}_i) | \boldsymbol{\delta}_i) \quad (9)$$

At the second stage, the hyperparameters are estimated through MLE by treating the estimated $\{\hat{\mathcal{M}}_i, i = 1, \dots, I\}$ at the first stage as observations. The duration hyperparameters $(\mu_d^{(s)}, \sigma_d^{2(s)}, s = 1, \dots, K)$ and drift rate hyperparameters $(\mu_0^{(s)}, \kappa_0^{2(s)}, s = 1, \dots, K + 1)$ can be obtained analytically as

$$\hat{\mu}_d^{(s)} = \frac{\sum_{i=1}^I \hat{d}_i^{(s)}}{I}, \quad \widehat{\sigma}_d^{2(s)} = \frac{\sum_{i=1}^I (\hat{d}_i^{(s)} - \hat{\mu}_d^{(s)})^2}{I} \quad (10)$$

and

$$\hat{\mu}_0^{(s)} = \frac{\sum_{i=1}^I \frac{\hat{\beta}_i^{(s)}}{\widehat{\sigma}_i^{2(s)}}}{\sum_{i=1}^I \frac{1}{\widehat{\sigma}_i^{2(s)}}}, \quad \widehat{\kappa}_0^{2(s)} = \frac{1}{I} \sum_{i=1}^I \frac{(\hat{\beta}_i^{(s)} - \hat{\mu}_0^{(s)})^2}{\widehat{\sigma}_i^{2(s)}} \quad (11)$$

For the hyperparameters ν_0, γ_0 in the inverse Gamma distribution, the closed form is not tractable and instead can be estimated through nonlinear optimization techniques.

In the above model specification and estimation, the critical parameter K needs to be selected first. AIC [36] or BIC [37] is typically used for model selection in regression. However, the conventional AIC or BIC is not very effective for multiple change-point models [38]. Besides, although increasing the number of change points may improve the model fitting accuracy, it may not necessarily result in better prognostic accuracy. In fact, increasing the number of change points will introduce extra uncertainties in RUL prediction (uncertainty of future change-point locations). Even if there is no over-fitting issue, it may still significantly reduce the prediction accuracy. Therefore there is tradeoff between model fitting accuracy and RUL predictability. To address this issue, we propose to use the cross validation technique for change-point model selection. Specifically, we apply leave-one-out-cross-validation approach. For each value K , the offline

model estimation and online RUL prediction are performed, and the average prediction error is calculated. Then the optimal K is the one with minimal average prediction error. Note that for linear regression models without any change points, the cross validation approach is asymptotically equivalent to BIC based model selection [20]. However, for change-point models or other general models, cross-validation would be better since it is directly based on the model predictability.

4 Online Model Updating and RUL Prediction

Once the prior information is calculated based on historical data at the offline stage, as described in the previous section, it can be utilized for the RUL prediction of a new in-service unit at the online stage. To do this accurately requires a sequential updating the posterior distributions of certain key parameters, such as the location of the latest change point, the number of change-point occurred, and the drift and diffusion parameter of the current phase, which is the main challenge. In this section, we will first show the details of how to update the model recursively, and then present the RUL prediction method based on the updated model.

4.1 Exact Bayesian Model Updating

Assume that we have observed the degradation signal up to the current time step m for an in-service unit, denoted as $X_{0:m} = (X_0, X_1, \dots, X_m)$. The objective of Bayesian model updating is to incorporate the new observations to the estimated model by computing the posterior distribution of model parameters. The target distribution that needs to be updated is $p(\boldsymbol{\theta}_m | X_{0:m})$ where state vector $\boldsymbol{\theta}_m = (\beta_m, \sigma_m^2, \tau_m, s_m)$. In general, the analytic expression for this joint posterior distribution is intractable. A natural way is to use sequential Monte Carlo method to get an approximation. Wen *et al.* [30] developed a stratified particle filtering algorithm for online model updating of a general path model. This method can effectively handle intractable posteriors. However, due to high dimensionality, at each time step a large number of particles have to be generated to guarantee the approximation accuracy. Therefore it is computationally expensive and may not be applicable for online monitoring. Fortunately, due to the assignment of conjugate priors for β and σ^2 , the posterior could be exactly calculated through a novel recursive updating approach. Given the observed data, the joint posterior distribution of all parameters can be derived as

$$p(\boldsymbol{\theta}_m | X_{0:m}) = p(\beta_m, \sigma_m^2, \tau_m, s_m | X_{0:m}) = P(\tau_m, s_m | X_{0:m}) p(\beta_m, \sigma_m^2 | \tau_m, s_m, X_{0:m}) \quad (12)$$

As we can see, the joint distribution can be factorized as the product of a posterior of the *discrete* components (τ_m and s_m), and the *continuous* components (β_m and σ_m^2). The discrete components are essential for phase tracking and future change-point prediction, while the continuous components are required to predict the degradation level at the end of the current phase. The details of how to calculate these two parts are given in the following paragraphs.

The conditional posterior distribution of the continuous components $p(\beta_m, \sigma_m^2 | \tau_m, s_m, X_{0:m})$ can be calculated based on Theorem 1 as follows.

Theorem 1. Given the conjugate priors shown in Eq. (4) for β and σ^2 , the conditional posterior pdf $p(\beta_m, \sigma_m^2 | \tau_m = j, s_m = s, X_{0:m})$, can be calculated as

$$p(\beta_m, \sigma_m^2 | \tau_m = j, s_m = s, X_{0:m}) = p(\beta_m | \sigma_m^2, \tau_m = j, s_m = s, X_{0:m}) p(\sigma_m^2 | \tau_m = j, s_m = s, X_{0:m}) \quad (13)$$

where

$$\begin{aligned} (\sigma_m^2 | \tau_m = j, s_m = s, X_{0:m}) &\sim IG\left(v_0^{(s)} + \frac{m-j}{2}, \gamma_0^{(s)} + \frac{H_{j+1,m}^{(s)}}{2}\right) \\ (\beta_m | \sigma_m^2, \tau_m = j, s_m = s, X_{0:m}) &\sim N(\mu_{j+1,m}^{(s)}, \sigma_m^2 \kappa_{j+1,m}^{2(s)}) \end{aligned} \quad (14)$$

and

$$\begin{aligned} H_{j+1,m}^{(s)} &= \left[\frac{\mu_0^{2(s)}}{\kappa_0^{2(s)}} + \sum_{i=j+1}^m \frac{\delta_i^2}{\lambda_i} - \left(\frac{\mu_0^{(s)}}{\kappa_0^{2(s)}} + \sum_{i=j+1}^m \delta_i \right)^2 \left(\sum_{i=j+1}^m \lambda_i + \frac{1}{\kappa_0^{2(s)}} \right)^{-1} \right] \\ \kappa_{j+1,m}^{2(s)} &= \left(\sum_{i=j+1}^m \lambda_i + \frac{1}{\kappa_0^{2(s)}} \right)^{-1} \\ \mu_{j+1,m}^{(s)} &= \left(\frac{\mu_0^{(s)}}{\kappa_0^{2(s)}} + \sum_{i=j+1}^m \delta_i \right) \kappa_{j+1,m}^{2(s)} \end{aligned} \quad (15)$$

The proof is given in Appendix A. For the discrete components, based on the Bayes' rule, the posterior can be derived as

$$P(\tau_m = j, s_m = s | X_{0:m}) \propto P(\tau_m = j, s_m = s | X_{0:m-1}) p(X_m | \tau_m = j, s_m = s, X_{0:m-1}) \quad (16)$$

In the above equation, $P(\tau_m = j, s_m = s | X_{0:m-1})$ is the predictive probability mass function (PMF), which can be recursively calculated by conditioning on the states of the previous time step:

$$P(\tau_m = j, s_m = s | X_{0:m-1}) = \sum_{j', s'} P(\tau_{m-1} = j', s_{m-1} = s' | X_{0:m-1}) P(\tau_m = j, s_m = s | \tau_{m-1} = j', s_{m-1} = s', X_{0:m-1}) \quad (17)$$

In Eq. (17), $P(\tau_{m-1} = j', s_{m-1} = s' | X_{0:m-1})$ is the posterior distribution of the discrete component obtained at the previous time step. $P(\tau_m = j, s_m = s | \tau_{m-1} = j', s_{m-1} = s', X_{0:m-1})$ is the prior state transition probability, which can be derived based on the Markov properties of the occurrence of the change points, i.e., the probability of the occurrence of a new change point or a new phase at the current time only depends on the duration and phase index at the previous time step. Based on the prior knowledge of the phase duration and total number of phases, the state transition probability can be obtained as

$$P(\tau_m = j, s_m = s | \tau_{m-1} = j', s_{m-1} = s', X_{0:m-1}) = \begin{cases} \frac{1 - G^{(s')}(t_m - t_{j'})}{1 - G^{(s')}(t_{m-1} - t_{j'})}, & \text{if } j = j' \text{ and } s = s' < K + 1 \\ 1, & \text{if } j = j' \text{ and } s = s' = K + 1 \\ \frac{G^{(s')}(t_m - t_{j'}) - G^{(s')}(t_{m-1} - t_{j'})}{1 - G^{(s')}(t_{m-1} - t_{j'})}, & \text{if } j = m - 1 \text{ and } s = s' + 1 \leq K + 1 \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

where $G^{(s')}(\cdot)$ is the cumulative distribution function of the s' th phase duration. The top probability refers to there not being a change-point between t_m and t_{m-1} and the current phase is not the last phase. The second probability refers to the current phase is already at the last phase. The third probability refers to the event that there is a change point occurring.

The predictive density $p(X_m | \tau_m = j, s_m = s, X_{0:m-1})$ in Eq. (16) can be calculated based on Theorem 2 as follows.

Theorem 2. Suppose the conjugate priors shown in Eq. (4) are assumed for β and σ^2 .

if $j < m - 1$, $(X_m | \tau_m = j, s_m = s, X_{0:m-1})$ follows a univariate t distribution given as

$$(X_m | \tau_m = j, s_m = s, X_{0:m-1}) \sim t_1 \left(2v_0^{(s)} + m - j - 1, \mu_{m|j+1:m-1}^{(s)}, \frac{2\gamma_{j+1,m-1}^{(s)} \eta_{m|j+1:m-1}^{(s)}}{2v_0^{(s)} + m - j - 1} \right) \quad (19)$$

where

$$\begin{aligned} \mu_{m|j+1:m-1}^{(s)} &= \mu_{j+1,m-1}^{(s)} \lambda_m + X_{m-1}, \\ \gamma_{j+1,m-1}^{(s)} &= \gamma_0^{(s)} + \frac{H_{j+1,m-1}^{(s)}}{2} \\ \eta_{m|j+1:m-1}^{(s)} &= \lambda_m + \kappa_{j+1,m-1}^{2(s)} \lambda_m^2 \end{aligned} \quad (20)$$

if $j = m - 1$, $(X_m | \tau_m = j, s_m = s, X_{0:m-1})$ follows a univariate t distribution given as

$$(X_m | \tau_m = j, s_m = s, X_{0:m-1}) \sim t_1 \left(2v_0^{(s)}, \mu_m^{(s)}, \frac{\gamma_0^{(s)} \eta_m^{(s)}}{v_0^{(s)}} \right) \quad (21)$$

where

$$\begin{aligned} \mu_m^{(s)} &= \mu_0^{(s)} \lambda_m + X_{m-1} \\ \eta_m^{(s)} &= \lambda_m + \kappa_0^{2(s)} \lambda_m^2 \end{aligned} \quad (22)$$

The proof is included in Appendix B. As we can see, based on Eq. (12)-(22), the posterior distributions of all parameters of interest can be exactly calculated through a recursive updating approach.

4.2 Controlling the Computational Cost

Compared with sequential Monte Carlo methods, the exact inference for sequential model updating runs much faster when m is not large. However, as m becomes very large, the algorithm may be very time-consuming. The reason is that it needs to calculate all the probabilities of $P(\tau_m = j, s_m = s | X_{0:m})$ for $j = 0, 1, \dots, m - 1$ and $s = 1, 2, \dots, K + 1$ at each time step m , which increases almost linearly with m . Therefore, the algorithm needs to be improved in a computationally efficient manner for real-time estimation. It is observed that as the number of observations increases, the PMF of the latest change point becomes more and more concentrated around the true change point, whereas the probabilities at other locations are close to zero. To reduce the computational load and balance the computational cost for all time steps, an approximation strategy with a set of support point of fixed size can be applied [21]. The basic idea is to approximate the posterior distribution of the discrete components by a small set of support points of a fixed size with significant probabilities. In other words, we set the posterior PMF to be

zero at those with negligible values and keep others with high probabilities. However, this strategy may result in zero PMFs for certain phases, i.e., $P(s_m = s|X_{0:m}) = 0$ or $P(\tau_m = j, s_m = s|X_{0:m}) = 0$ for all $j = 0, 1, \dots, m - 1$, whose true values may be just temporally small and will become significant as more observations are obtained. To avoid this situation, we propose to use a stratified sampling method, where for each s , N locations with highest PMF are selected for the latest change points. The algorithm is summarized in Algorithm 1. With this strategy, the maximum computational cost for each time step m can be controlled effectively.

Algorithm 1 The Fixed Support Size Strategy for Model Updating

For each time step m :

If $m \leq N - 1$:

- Calculate $P(\tau_m = j, s_m = s|X_{0:m})$ for $j = 0, 1, m - 1, s = 1, \dots, K + 1$.

If $m > N$:

- For each s , select $N - 1$ support points with highest probabilities $P(\tau_{m-1} = j, s_{m-1} = s|X_{0:m-1})$
 - Calculate $P(\tau_m = j, s_m = s|X_{0:m})$ at all these selected $(N - 1)$ support points and at the current time step m for $s = 1, \dots, K + 1$, set others to be zero.
 - Normalize the probabilities.
-

4.3 RUL prediction

Once the parameters in the model have been updated, the RUL of the in-service unit can be predicted. A failure is typically defined as the event that the degradation signal first hits the failure threshold Γ . Denote the RUL at the current time t_m as R_m . Based on the concept of first passage time (FPT), the RUL can be formulated as $R_m = \inf\{l: X(t_m + l) \geq \Gamma | X_{0:m}\}$. For the conventional Wiener process degradation model, RUL has been proven to follow an inverse Gaussian distribution. For an in-service unit with observations $X_{0:m}$, if β and σ^2 are fixed, the pdf of the residual life can be derived as [39]

$$f_{R_m}(l|X_m, \beta, \sigma^2) = \frac{\Gamma - X_m}{\sqrt{2\pi\sigma^2 l^3}} \exp\left(-\frac{(\Gamma - X_m - \beta l)^2}{2\sigma^2 l}\right) \quad (23)$$

However, in our model, due to the unknown change points and randomness of β, σ^2 at each phase, the RUL is very complicated and intractable analytically. For model consistency, we assume that

the degradation amplitude will not exceed the failure threshold before the last change point. In RUL prediction, we need to first predict the location of the final change point t_{fc} and the degradation amplitude $X(t_{fc})$, and then conditioning on them to predict when the last phase will hit the failure threshold. Mathematically, the pdf of RUL can be represented as

$$f_{R_m}(l|X_{0:m}) = \iint p(t_{fc}, X(t_{fc})|X_{0:m})f_{R_m}(l|t_{fc}, X(t_{fc}), X_{0:m})dX(t_{fc})dt_{fc} \quad (24)$$

where

$$\begin{aligned} p(t_{fc}, X(t_{fc})|X_{0:m}) &= \sum_{s_m, \tau_m} P(s_m, \tau_m|X_{0:m})p(t_{fc}, X(t_{fc})|s_m, \tau_m, X_{0:m}) \\ &= \sum_{s_m, \tau_m} P(s_m, \tau_m|X_{0:m}) \int p(t_{fc}, X(t_{fc})|\boldsymbol{\theta}_m)p(\beta_m, \sigma_m^2|s_m, \tau_m, X_{0:m})d\beta_m d\sigma_m^2 \end{aligned} \quad (25)$$

and

$$\begin{aligned} f_{R_m}(l|t_{fc}, X(t_{fc}), X_{0:m}) \\ = \iint f_{R_m}(l|t_{fc}, X(t_{fc}), X_{0:m}, \beta^{(K+1)}, \sigma^{2(K+1)})p(\beta^{(K+1)}, \sigma^{2(K+1)}|t_{fc}, X(t_{fc}), X_{0:m})d\beta^{(K+1)}d\sigma^{2(K+1)} \end{aligned} \quad (26)$$

It is worth noting that here t_{fc} may be less than t_m , therefore $f_{R_m}(l|t_{fc}, X(t_{fc}), X_{0:m})$ may depend on $X_{0:m}$ or may be independent of $X_{0:m}$. Clearly, the RUL is intractable due to multiple complex integrations. A natural way to address this issue is to use Monte Carlo simulation approach. Specifically, we can first generate M samples for the current state vector $\boldsymbol{\theta}_m$ through the updated posterior distribution $p(\boldsymbol{\theta}_m|X_{0:m})$, and then conditioning on each sample, simulate the remaining change points, β and σ^2 of each phase, and degradation levels at remaining change points. The duration of the last phase can be directly sampled from inverse Gaussian distribution based on Eq. (23). The details of the Monte Carlo approach for RUL sampling ($T_i, i = 1, \dots, M$) is illustrated in Algorithm 2. It should be mentioned that for multiple change-point based model, the probability of being at the last phase at early prediction stage is inevitably nonzero, i.e., $P(s_m = K + 1|X_{0:m}, \tau_m = j) > 0$ while the actual phase $s_m < K + 1$. This probability may even be significant for certain signals due to inherent randomness of Wiener process. If the actual degradation rate is very small at the current time, the sampled failure time may be significantly larger than the actual value. Besides, for phases before the last phase, negative drift parameters may be sampled, which may also significantly increase the residual life. To make it more accurate, we apply lower bounds $l^{(s)}$ for all the drift parameters as constraints in the sampling process,

which can be obtained by selecting the minimal drift parameter of each phase of all historical signals.

Algorithm 2 Monte Carlo Simulation for RUL Prediction

Generate samples from $p(\theta_m | X_{0:m})$:

For $i = 1:M$

 Sample $(\tau_{m,i}, s_{m,i}) \sim P(\tau_m, s_m | X_{0:m})$ based on Eq. (16)

 Sample $(\beta_{m,i}, \sigma_{m,i}^2) \sim p(\beta_m, \sigma_m^2 | \tau_{m,i}, s_{m,i}, X_{0:m})$ based on Eq. (14)

End

Simulate RUL

For $i = 1:M$

 If $s_{m,i} = K + 1$, then

 Sample $T_i \sim f_{R_m}(l | X_m, \beta_{m,i}, \sigma_{m,i}^2)$ based on Eq. (23)

 else

 For $s = s_{m,i}: K$

 If $s = s_{m,i}$

 Sample $t_{c_{s,i}} \sim t_{c_{s-1,i}} + N(d | \mu_d^{(s)}, \sigma_d^{2(s)}, d \geq t_m - t_{\tau_{m,i}})$

 Sample $X(t_{c_{s,i}}) \sim X(t_m) + N(X | \beta_{m,i}(t_{c_{s,i}} - t_m), (t_{c_{s,i}} - t_m)\sigma_{m,i}^2)$

 else

 Sample $t_{c_{s,i}} \sim t_{c_{s-1,i}} + N(d | \mu_d^{(s)}, \sigma_d^{2(s)})$

 Sample $(\beta_i^{(s)}, \sigma_i^{2(s)}) \sim IG(\nu_0^{(s)}, \gamma_0^{(s)}) N(\mu_0^{(s)}, \sigma_0^{2(s)} \kappa_0^{2(s)})$

 Sample $X(t_{c_{s,i}}) \sim X(t_{c_{s-1,i}}) + N(X | \beta_i^{(s)}(t_{c_{s,i}} - t_{c_{s-1,i}}), (t_{c_{s,i}} - t_{c_{s-1,i}})\sigma_i^{2(s)})$

 End

 End

 Sample $(\beta_i^{(K+1)}, \sigma_i^{2(K+1)}) \sim IG(\nu_0^{(K+1)}, \gamma_0^{(K+1)}) N(\mu_0^{(K+1)}, \sigma_0^{2(K+1)} \kappa_0^{2(K+1)})$

 Sample $L_i \sim f_{R_{c_K}}(l | X(t_{c_{K,i}}), \beta_i^{(K+1)}, \sigma_i^{2(K+1)})$ based on Eq. (23)

$T_i = (t_{c_{K,i}} - t_m) + L_i$

 End

End

5 Case Studies

In this section, we first use simulation study to demonstrate the robustness and effectiveness of our model. Then, the proposed model is applied to real case study of rotational bearings.

5.1 Simulation Study

In this subsection, the prediction is illustrated and the performance is evaluated through numerical simulations. For simplicity, we only consider one-change-point and two-change-point scenarios in the simulation model. The hyperparameters for each scenario are specified in Table 1. The failure

threshold is set to be $\Gamma = 20$. The inspection interval is set to be $\Delta t = 2$. For each scenario, 200 degradation signals are generated. The support size for the PMF approximation at the updating stage is specified as $N = 20$. The number of samples for RUL prediction is set as $M = 3000$. Due to the randomness of Monte Carlo method, the prediction procedure is repeated 10 times for each signal.

Table 1. Hyperparameters for simulation

Variables	$K_{\text{true}} = 1$	$K_{\text{true}} = 2$
$d^{(s)}$	$\mu_d^{(1)} = 300, \sigma_d^{2(1)} = 10^2$ $\mu_d^{(2)} = 400, \sigma_d^{2(2)} = 10^2$	$\mu_d^{(1)} = 200, \sigma_d^{2(1)} = 10^2$ $\mu_d^{(2)} = 300, \sigma_d^{2(2)} = 10^2$ $\mu_d^{(3)} = 500, \sigma_d^{2(3)} = 10^2$
$\beta^{(s)}$	$\mu_0^{(1)} = 0.01, \kappa_0^{2(1)} = 0.008$ $\mu_0^{(2)} = 0.06, \kappa_0^{2(2)} = 0.006$	$\mu_0^{(1)} = 1 \times 10^{-5}, \kappa_0^{2(1)} = 9 \times 10^{-4}$ $\mu_0^{(2)} = 0.02, \kappa_0^{2(2)} = 0.002$ $\mu_0^{(3)} = 0.09, \kappa_0^{2(3)} = 0.002$
$\sigma^{2(s)}$	$v_0^{(1)} = 2, \gamma_0^{(1)} = 0.05$ $v_0^{(2)} = 2, \gamma_0^{(2)} = 0.1$	$v_0^{(1)} = 2, \gamma_0^{(1)} = 0.06$ $v_0^{(2)} = 2, \gamma_0^{(2)} = 0.08$ $v_0^{(3)} = 2, \gamma_0^{(3)} = 0.1$

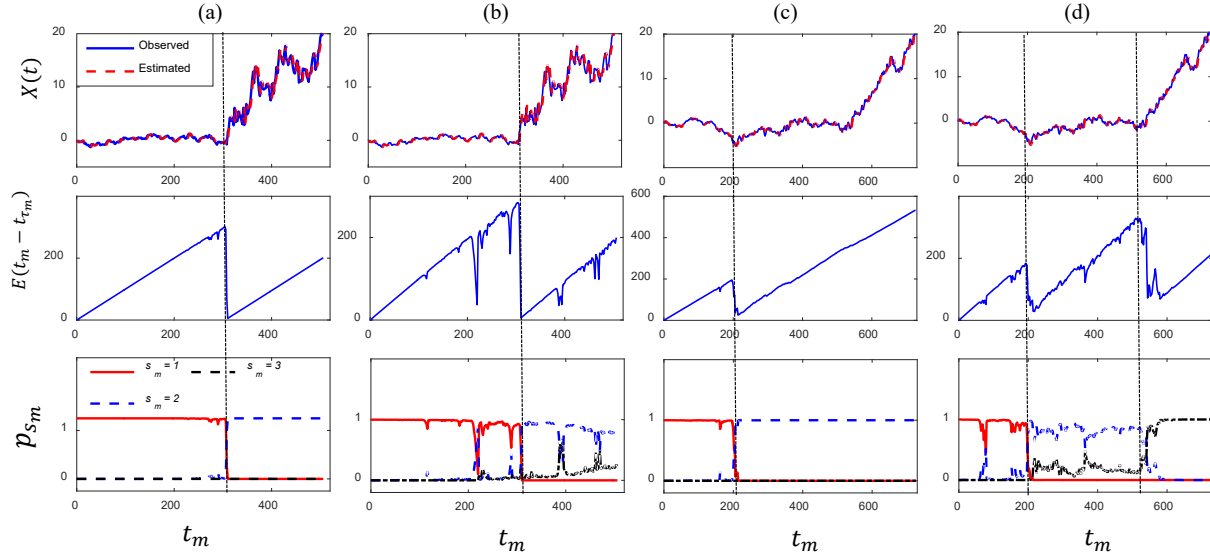


Figure 2. Illustration of the online model updating process. (a) $K_{\text{true}} = 1, K = 1$; (b) $K_{\text{true}} = 1, K = 2$ (c) $K_{\text{true}} = 2, K = 1$ and (d) $K_{\text{true}} = 2, K = 2$; top panel: observed and filtered signals; middle panel: the expected duration of the current phase; bottom panel: the posterior PMF of the index of current phase. The vertical dashed lines are true change-point locations.

Due to space limitation, the estimated hyperparameters are not provided here. Using the leave-one-out-cross-validation approach, we find that the identified optimal change-point number for each

dataset is equivalent to the true value, which demonstrates the effectiveness of the proposed approach for model selection (see Figure 5). Figure 2 shows the online model updating process for two degradation signals, each of which is randomly selected from each dataset. The first and the second column shows the results by assuming $K = 1$ and $K = 2$, respectively, while the true value $K_{true} = 1$. Similarly, the third and fourth column assume $K = 1$ and $K = 2$, respectively, while $K_{true} = 2$. Here we select two K 's for each signal to study the consequence if K is specified inappropriately. Clearly, if an appropriate K is specified (Figure 2 (a) and (d)), the recursive updating algorithm can accurately detect the occurrence of change points and track the phase index. However, if K is larger (or less) than the actual value, more phases will occur (or some change points may not be detected), which will thus affect the model fitting and RUL prediction.

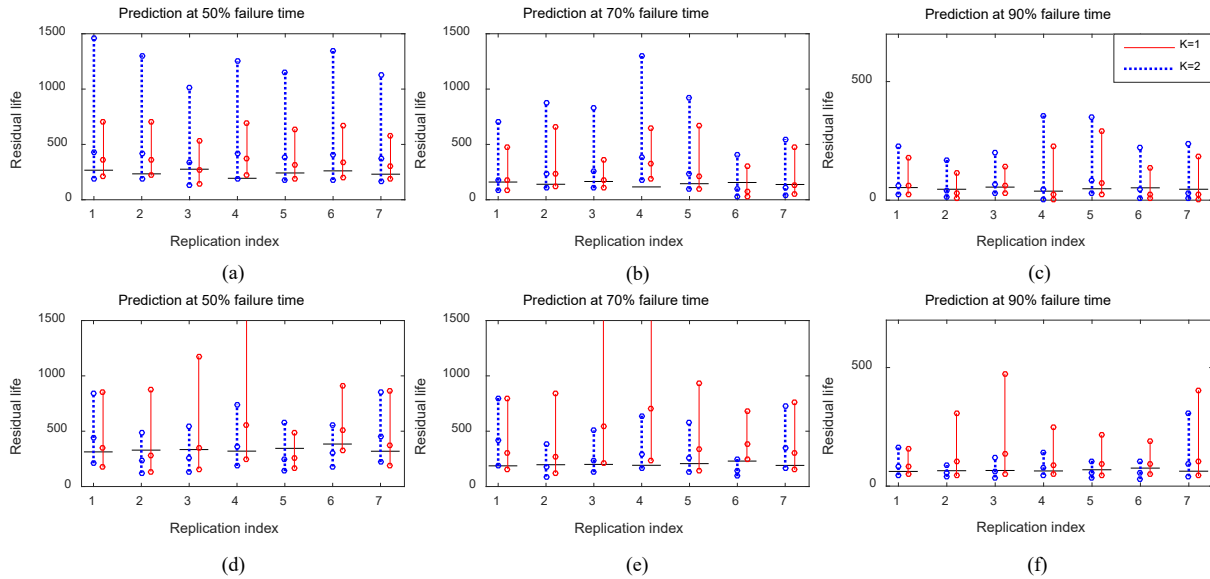


Figure 3. Prediction intervals of RUL for 14 simulated signals predicting at 50%, 70%, 90% of failure time. (a-c): $K_{true} = 1$; (d-f): $K_{true} = 2$. The "o" denotes the 5%, 50% and 95% quantile of RUL prediction; "—" represent the actual failure time.

Figure 3 shows the prediction intervals of 7 randomly selected signals from each dataset at three different prediction times, i.e., 50%, 70%, 90% of actual failure time. Figure 4 illustrates the pdf of RUL of the third and the second unit of the seven signals in each category. To compare the prediction performance of different models, we select $K = 1$ and $K = 2$ for both two types of signals. As expected, for both model specifications, the prediction becomes more and more accurate for all signals as more and more observations are collected. This characteristic is highly desirable since it becomes more important to get an accurate RUL prediction as the unit approaches

failure. Comparing the two model selections, we can see that if the true K is selected, the prediction performance is much better than if K is selected inappropriately. Note that in the simulation, the right model was effectively selected through the cross-validation approach.

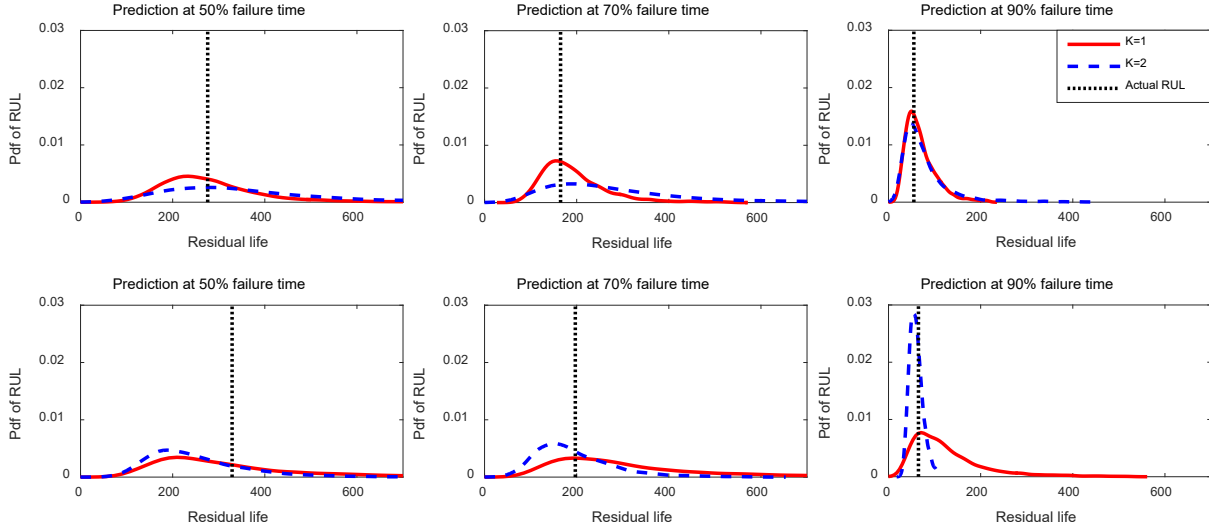


Figure 4. Comparison of the RUL prediction between $K = 1$ and $K = 2$. Top panel (a-c): $K_{true} = 1$; bottom panel (d-f): $K_{true} = 2$.

To further evaluate and analyze the overall prediction accuracy, 200 signals in each category are used for testing. We use the root-mean-square deviation (RMSD) as a performance metric, which is defined as

$$\text{RMSD} = \sqrt{\frac{1}{I} \sum_{i=1}^I E(\hat{R}_i - R_{i,\text{true}})^2} \quad (27)$$

where I is the total number of units, \hat{R}_i and $R_{i,\text{true}}$ are the predicted and true RUL of unit i , respectively. Figure 5 shows the RMSD at six prediction times for both two types of signals. For comparison purpose, three models $K = 0, 1, 2$ are provided here. Clearly, the model with appropriate K outperforms all other models at all prediction times. From Figure 5(a) we can see that, the RMSD goes down gradually for both $K = 1$ and $K = 2$ models, indicating that with more observations collected, the prediction becomes more accurate. However, the model with $K = 1$ is more accurate in prediction than $K = 2$. The reason is that adding excessive change points will introduce unnecessary uncertainty in RUL prediction, e.g., uncertainty regarding future change points, and Wiener process parameters of the last phase.

For $K = 0$, the RMSD first increases and then decreases. Similarly phenomenon can also be observed in Figure 5(b). If the number of change points is selected insufficiently, the RMSD first increases and then decreases as we increase the prediction times. The reason is that the early phases often have smaller degradation rates than the later phases, as shown in Figure 2. If insufficient change points are assumed, the early phases may be mistakenly detected as the last phase. The more observations are collected, the less influence the prior will have on the posterior updating and thus the lower the updated degradation rate will be. Consequently it will make the predicted RUL much higher than the actual value. When the unit approaches to its failure, the prediction accuracy for all of these models increases. Therefore, the parameter K is critical for the prognostic model to produce an accurate prediction.

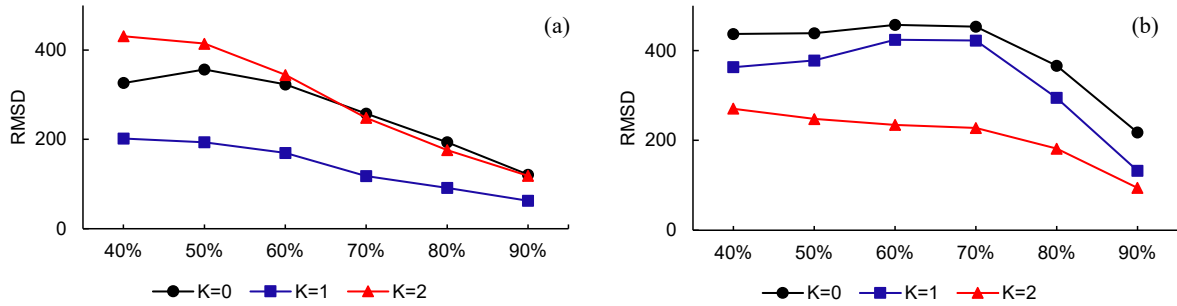


Figure 5. Comparison of the RMSD at six prediction times: (a) $K_{true} = 1$; (b) $K_{true} = 2$.

5.2 Application to Bearing Signals

In this subsection, the proposed approach is applied to the degradation signal of rotational bearings. These degradation signals are log-transformed vibrational signals obtained through an accelerated testing on a set of identical thrust ball bearings [3, 26]. In total, there are 25 degradation signals. The time interval for inspection is 2 minutes, i.e., $\Delta t = 2$ for all units. Figure 6 shows three degradation paths for illustration. The failure threshold is specified as $\Gamma = \ln(0.03)$, which is based on the published industrial standards [3]. There are obviously two phases for all signals. Moreover, the locations of change points vary from unit to unit.

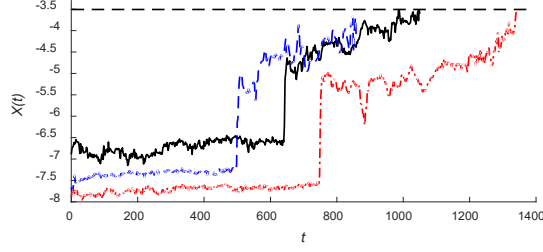


Figure 6. Degradation paths of three representative bearings.

Table 2. Estimated hyperparameters for three models

Variables	$K = 0$	$K = 1$	$K = 2$
$d(s)$	$\mu_d^{(1)} = 676$ $\sigma_d^{2(1)} = 255^2$	$\mu_d^{(1)} = 360, \sigma_d^{2(1)} = 241^2$ $\mu_d^{(2)} = 317, \sigma_d^{2(2)} = 134^2$	$\mu_d^{(1)} = 310, \sigma_d^{2(1)} = 190^2$ $\mu_d^{(2)} = 62, \sigma_d^{2(2)} = 119^2$ $\mu_d^{(3)} = 306, \sigma_d^{2(3)} = 142^2$
$\beta(s)$	$\mu_0^{(1)} = 0.0056$ $\kappa_0^{2(1)} = 8.23 \times 10^{-4}$	$\mu_0^{(1)} = 8.317 \times 10^{-4}$ $\kappa_0^{2(1)} = 0.0038$ $\mu_0^{(2)} = 0.0083$ $\kappa_0^{2(2)} = 0.0012$	$\mu_0^{(1)} = -3.12 \times 10^{-4}, \kappa_0^{2(1)} = 0.006$ $\mu_0^{(2)} = 0.0215, \kappa_0^{2(2)} = 0.4425$ $\mu_0^{(3)} = 0.0049, \kappa_0^{2(3)} = 0.0024$
$\sigma^{2(s)}$	$v_0^{(1)} = 10.07$ $\gamma_0^{(1)} = 0.067$	$v_0^{(1)} = 1.39, \gamma_0^{(1)} = 0.003$ $v_0^{(2)} = 2.29, \gamma_0^{(2)} = 0.017$	$v_0^{(1)} = 2.1, \gamma_0^{(1)} = 0.004$ $v_0^{(2)} = 0.42, \gamma_0^{(2)} = 0.004$ $v_0^{(3)} = 5.1, \gamma_0^{(3)} = 0.03$

For model selection and comparison, we mainly consider three cases, $K = 0, 1,$ and 2 . Table 2 summarizes the estimated hyperparameters for each case. From Table 2 we can see that, if $K=1$ is specified, the second phase has a larger drift rate than the first one, which indicates a faster degradation. If $K = 2$ is specified, the degradation rate of the second phase is much larger than the third phase. Not surprisingly, almost all bearings have a sudden jump after the first phase. Although extremely short in duration, the sudden jump is so significant that it was identified as a single phase at the offline model fitting for all signals. Even for online model updating, as shown in Figure 7(b), the sudden jump is detected as a single phase.

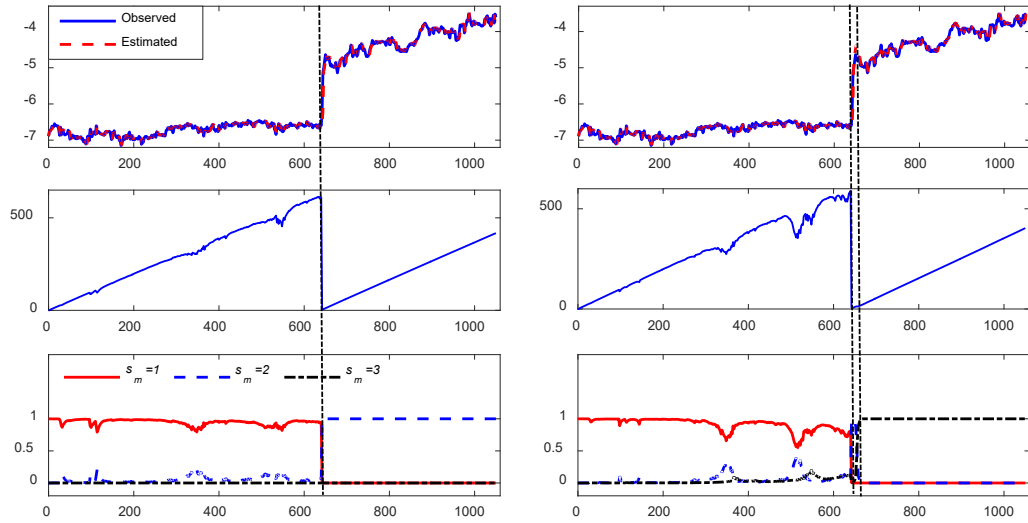


Figure 7. Illustration of online model updating process. (a): $K = 1$; (b): $K = 2$; top panel: observed and filtered signals; middle panel: the expected duration of the current phase; bottom panel: the posterior PMF of the index of current segment. The vertical dashed lines are change-point locations identified at the offline stage.

Figure 8 shows the RMSD of the 25 bearing signals at three prediction times for three models. Figure 9 shows the overall prediction accuracy at six prediction times. Clearly, at all the six prediction times, the model with only one change point performs best in terms of the overall RMSD. As observed in our study, specifying two change points only improves the prediction within the sudden-jump phase. It often reduce the prediction accuracy at the early stage and even at the late stage. Besides, assuming no change points is surprisingly better than using two and three change points. The reason is that for some signals, there is a sharp increase right before failure, as can be seen from the red curve in Figure 6. For those signals, specifying two more change points may improve individual model fitting. However, the large uncertainty of change-point locations and degradation rate dominates the prediction accuracy and thus the performance is reduced.

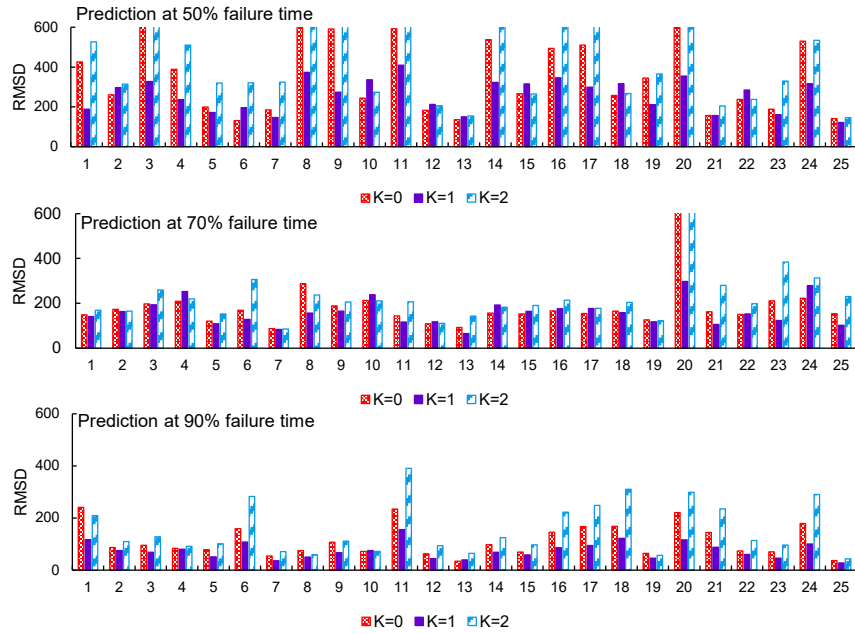


Figure 8. Comparison of RMSD for three models.

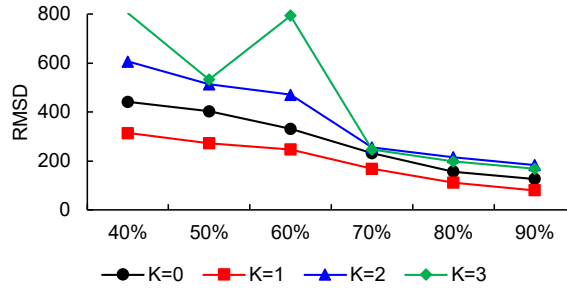


Figure 9. Prediction accuracy at different times for $K = 0, 1$ and 2

6 Conclusion and Discussion

In this paper, a Bayesian multiple change-point Wiener process is proposed for degradation modeling and online RUL prediction. To take into account the unit heterogeneity, all the model parameters except the number of change points are modeled with random distributions. At the offline stage, an empirical two-stage process is proposed for model estimation. Besides, a cross-validation approach is proposed for optimal change-point number selection. At the online monitoring stage, an exact recursive updating method is developed to sequentially calculate the joint posterior distribution of key parameters, including the latest change point, phase index, and model parameters of the current phase, which is essential for RUL prediction. To control the

computational cost, a fixed-support-size strategy is proposed, which can effectively control and balance the computational load of each time step yet without influencing the estimation accuracy. In RUL prediction, an effective Monte Carlo simulation algorithm is proposed. Simulation and real case studies demonstrate that the proposed prognostic framework can effectively improve the prediction accuracy.

There are still some issues that are worthy of further investigation. First, in the proposed method, the change points are assumed to be exactly on the inspection epochs in the offline model estimation and online model updating. In practice, the change points can be at any location between inspection epochs. Releasing this constraint may improve the model accuracy. In addition, in the current method, a linear drift function is assumed for all phases. In practice, however, nonlinear drift functions or a mixture of linear and nonlinear drift functions may be more preferable. It may significantly reduce the model complexity and consequently the model uncertainty. Furthermore, other stochastic processes, such as inverse Gaussian process, and Gamma process, may model the degradation signal better. How to incorporate change points into these stochastic processes needs to be investigated.

Appendix A: Proof of Theorem 1

For notational convenience, we ignore the subscript m for β_m and σ_m^2 , and ignore the subscript s for phase index. Suppose $\pi(\beta, \sigma^2) = IG(v_0, \gamma_0)N(\mu_0, \sigma^2 \kappa_0^2)$

$$\begin{aligned}
p(\beta, \sigma^2 | \delta_{j+1:m}, \tau_m = j) &\propto p(\beta, \sigma^2) p(\delta_{j+1:m} | \beta, \sigma^2) \\
&\propto \left[\frac{1}{(2\pi\sigma^2\kappa_0^2)^{\frac{1}{2}}} e^{-\frac{(\beta-\mu_0)^2}{2\sigma^2\kappa_0^2}} \right] \left[\frac{\gamma_0^{v_0}}{\Gamma(v_0)} (\sigma^2)^{-v_0-1} e^{-\frac{\gamma_0}{\sigma^2}} \right] \left[(2\pi)^{-\frac{m-j}{2}} (\sigma^2)^{-\frac{m-j}{2}} e^{-\frac{\sum_{i=j+1}^m (\delta_i - \beta\lambda_i)^2 / \lambda_i}{2\sigma^2}} \right] \\
&\propto (\sigma^2)^{-v_0-1-\frac{m-j}{2}} \frac{1}{2\pi(\sigma^2)^{1/2}} \exp \left[-\frac{(\beta - \mu_0)^2 / \kappa_0^2 + 2\gamma_0 + \sum_{i=j+1}^m (\delta_i - \beta\lambda_i)^2 / \lambda_i}{2\sigma^2} \right] \\
&\propto (\sigma^2)^{-v_0-1-\frac{m-j}{2}} \frac{1}{2\pi(\sigma^2)^{\frac{1}{2}}} \exp \left(-\frac{\beta^2 \left(\sum_{i=j+1}^m \lambda_i + \frac{1}{\kappa_0^2} \right) - 2\beta \left(\frac{\mu_0}{\kappa_0^2} + \sum_{i=j+1}^m \delta_i \right) + \frac{\mu_0^2}{\kappa_0^2} + 2\gamma_0 + \sum_{i=j+1}^m \frac{\delta_i^2}{\lambda_i}}{2\sigma^2} \right) \\
&\propto (\sigma^2)^{-v_0-1-\frac{m-j}{2}} \frac{1}{2\pi(\sigma^2)^{\frac{1}{2}}} \exp \left(-\frac{(\beta - \mu_m)^2}{2\sigma^2 \left(\sum_{i=1}^m \lambda_{j+i} + \frac{1}{\kappa_0^2} \right)^{-1}} \right)
\end{aligned}$$

$$\exp\left(-\frac{\frac{\mu_0^2}{\kappa_0^2} + 2\gamma_0 + \sum_{i=j+1}^m \frac{\delta_i^2}{\lambda_i} - \left(\frac{\mu_0}{\kappa_0^2} + \sum_{i=j+1}^m \delta_i\right)^2 \left(\sum_{i=j+1}^m \lambda_i + \frac{1}{\kappa_0^2}\right)^{-1}}{2\sigma^2}\right)$$

$$\propto IG(v_{j+1,m}, \gamma_{j+1,m}) N(\mu_{j+1,m}, \sigma^2 \kappa_{j+1,m}^2)$$

where

$$v_{j+1,m} = v_0 + \frac{m-j}{2}$$

$$\gamma_{j+1,m} = \gamma_0 + \frac{\left[\frac{\mu_0^2}{\kappa_0^2} + \sum_{i=j+1}^m \frac{\delta_i^2}{\lambda_i} - \left(\frac{\mu_0}{\kappa_0^2} + \sum_{i=j+1}^m \delta_i\right)^2 \left(\sum_{i=j+1}^m \lambda_i + \frac{1}{\kappa_0^2}\right)^{-1}\right]}{2}$$

$$\mu_{j+1,m} = \left(\frac{\mu_0}{\kappa_0^2} + \sum_{i=j+1}^m \delta_i\right) \left(\sum_{i=j+1}^m \lambda_i + \frac{1}{\kappa_0^2}\right)^{-1}$$

$$\kappa_{j+1,m}^2 = \left(\sum_{i=j+1}^m \lambda_i + \frac{1}{\kappa_0^2}\right)^{-1}$$

Appendix B: Proof of Theorem 2

For notational convenience, we ignore the subscript s for phase index. Based on Wiener process increment property

$$X_m = X_{m-1} + \beta_m \lambda_m + \sigma_m B(\lambda_m)$$

Note that here β_m and σ_m are the drift and diffusion parameters from t_{τ_m} to t_m .

If $\tau_m = j < m - 1$, then $\tau_{m-1} = j$, $\beta_m = \beta_{m-1}$, $\sigma_m = \sigma_{m-1}$. Based on Theorem 1 we can get

$$(\beta_{m-1} | \sigma_{m-1}^2, \tau_{m-1} = j, s_{m-1} = s, X_{0:m-1}) \sim N(\mu_{j+1,m-1}, \sigma_{m-1}^2 \kappa_{j+1,m-1}^2)$$

$$(\sigma_{m-1}^2 | \tau_{m-1} = j, s_{m-1} = s, X_{0:m-1}) \sim IG(v_{j+1,m-1}, \gamma_{j+1,m-1})$$

Then

$$(X_m | \sigma_{m-1}^2, \tau_m = \tau_{m-1} = j, s_{m-1} = s, X_{0:m-1}) \sim N(\mu_{j+1,m-1} \lambda_m + X_{m-1}, \sigma_{m-1}^2 (\lambda_m + \kappa_{j+1,m-1}^2 \lambda_m^2))$$

Let $\mu' = \mu_{j+1,m-1} \lambda_m + X_{m-1}$, $\eta' = \lambda_m + \kappa_{j+1,m-1}^2 \lambda_m^2$, then

$$p(X_m | \tau_m = j, s_m = s, X_{0:m-1})$$

$$\propto \int_0^\infty (\sigma_{m-1}^2 \eta')^{-\frac{1}{2}} e^{\left[\frac{-(X_m - \mu')^2}{2\sigma_{m-1}^2 \eta'}\right]} (\sigma_{m-1}^2)^{-v_0 - \frac{m-1-j}{2} - 1} e^{-\frac{\gamma_{j+1,m-1}}{\sigma_{m-1}^2}} d\sigma_{m-1}^2$$

$$\begin{aligned}
& \propto (\sigma_{m-1}^2)^{-v_0 - \frac{m-j}{2} - 1} e^{-\left[\frac{(X_m - \mu')^2 + 2\gamma_{j+1, m-1} \eta'}{2\sigma_{m-1}^2 \eta'} \right]} d\sigma_{m-1}^2 \\
& \propto \frac{\Gamma(v_0 + \frac{m-j}{2})}{\left[\frac{(X_m - \mu')^2}{2\eta'} + \gamma_{j+1, m-1} \right]^{v_0 + \frac{m-j}{2}}} \\
& \propto \left[1 + \frac{1}{2v_0 + m - j - 1} \frac{(X_m - \mu')^2 (2v_0 + m - j - 1)}{2\gamma_{j+1, m-1} \eta'} \right]^{-\frac{(2v_0 + m - j)}{2}} \\
& (X_m | X_{0:m-1}, \tau_m = j) \sim t_1 \left(2v_0 + m - j - 1, \mu', \frac{2\gamma_{j+1, m-1} \eta'}{2v_0 + m - j - 1} \right)
\end{aligned}$$

If $j = m - 1$, similarly, we can get

$$(X_m | X_{0:m-1}, \tau_m = m - 1) \sim t_1 \left(2v_0, \mu', \frac{\gamma_0 \eta'}{v_0} \right)$$

where $\mu' = \mu_0 \lambda_m + X_{m-1}$, $\eta' = \lambda_m + \kappa_0^2 \lambda_m^2$

References

- [1] L. Bian, N. Gebraeel, and J. P. Kharoufeh, "Degradation modeling for real-time estimation of residual lifetimes in dynamic environments," *IIE Transactions*, vol. 47, pp. 471-486, 2015.
- [2] W. B. Nelson, *Accelerated testing: statistical models, test plans, and data analysis* vol. 344: John Wiley & Sons, 2009.
- [3] N. Z. Gebraeel, M. A. Lawley, R. Li, and J. K. Ryan, "Residual-life distributions from component degradation signals: A Bayesian approach," *IIE Transactions*, vol. 37, pp. 543-557, 2005.
- [4] F. Zhao, Z. Tian, and Y. Zeng, "Uncertainty quantification in gear remaining useful life prediction through an integrated prognostics method," *IEEE Transactions on Reliability*, vol. 62, pp. 146-159, 2013.
- [5] S. T. Tseng, J. Tang, and I. H. Ku, "Determination of burn-in parameters and residual life for highly reliable products," *Naval Research Logistics (NRL)*, vol. 50, pp. 1-14, 2003.
- [6] R. R. Zhou, N. Serban, and N. Gebraeel, "Degradation modeling applied to residual lifetime prediction using functional data analysis," *The Annals of Applied Statistics*, pp. 1586-1610, 2011.
- [7] F. Camci and R. B. Chinnam, "Health-state estimation and prognostics in machining processes," *IEEE Transactions on automation science and engineering*, vol. 7, pp. 581-597, 2010.
- [8] J. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," *Mechanical systems and signal processing*, vol. 25, pp. 1803-1836, 2011.
- [9] A. Heng, A. C. Tan, J. Mathew, N. Montgomery, D. Banjevic, and A. K. Jardine, "Intelligent condition-based prediction of machinery reliability," *Mechanical systems and signal processing*, vol. 23, pp. 1600-1614, 2009.
- [10] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical systems and signal processing*, vol. 20, pp. 1483-1510, 2006.

- [11] G. J. Vachtsevanos, F. Lewis, A. Hess, and B. Wu, *Intelligent fault diagnosis and prognosis for engineering systems*: Wiley Online Library, 2006.
- [12] Y. Peng, M. Dong, and M. J. Zuo, "Current status of machine prognostics in condition-based maintenance: a review," *The International Journal of Advanced Manufacturing Technology*, vol. 50, pp. 297-313, 2010.
- [13] S.-J. Wu and C.-T. Chang, "Optimal design of degradation tests in presence of cost constraint," *Reliability Engineering & System Safety*, vol. 76, pp. 109-115, 2002.
- [14] H. Lim and B.-J. Yum, "Optimal design of accelerated degradation tests based on Wiener process models," *Journal of Applied Statistics*, vol. 38, pp. 309-325, 2011.
- [15] C. J. Lu and W. O. Meeker, "Using degradation measures to estimate a time-to-failure distribution," *Technometrics*, vol. 35, pp. 161-174, 1993.
- [16] M. A. Freitas, M. L. G. de Toledo, E. A. Colosimo, and M. C. Pires, "Using degradation data to assess reliability: a case study on train wheel degradation," *Quality and Reliability Engineering International*, vol. 25, pp. 607-629, 2009.
- [17] Z. S. Ye and M. Xie, "Stochastic modelling and analysis of degradation for highly reliable products," *Applied Stochastic Models in Business and Industry*, vol. 31, pp. 16-32, 2015.
- [18] M. Pandey, X.-X. Yuan, and J. Van Noortwijk, "The influence of temporal uncertainty of deterioration on life-cycle management of structures," *Structure and Infrastructure Engineering*, vol. 5, pp. 145-156, 2009.
- [19] Z. Pan and N. Balakrishnan, "Reliability modeling of degradation of products with multiple performance characteristics based on gamma processes," *Reliability Engineering & System Safety*, vol. 96, pp. 949-957, 2011.
- [20] Z.-S. Ye and N. Chen, "The inverse Gaussian process as a degradation model," *Technometrics*, vol. 56, pp. 302-311, 2014.
- [21] Z. Ye, N. Chen, and K. L. Tsui, "A Bayesian approach to condition monitoring with imperfect inspections," *Quality and Reliability Engineering International*, vol. 31, pp. 513-522, 2015.
- [22] G. Whitmore and F. Schenkelberg, "Modelling accelerated degradation data using Wiener diffusion with a time scale transformation," *Lifetime data analysis*, vol. 3, pp. 27-45, 1997.
- [23] X.-S. Si, W. Wang, C.-H. Hu, D.-H. Zhou, and M. G. Pecht, "Remaining useful life estimation based on a nonlinear diffusion degradation process," *IEEE Transactions on Reliability*, vol. 61, pp. 50-67, 2012.
- [24] D. R. Cox and H. D. Miller, *The theory of stochastic processes* vol. 134: CRC Press, 1977.
- [25] J. Son, Y. Zhang, C. Sankavaram, and S. Zhou, "RUL prediction for individual units based on condition monitoring signals with a change point," *IEEE Transactions on Reliability*, vol. 64, pp. 182-196, 2015.
- [26] N. Chen and K. L. Tsui, "Condition monitoring and remaining useful life prediction using degradation signals: Revisited," *IIE Transactions*, vol. 45, pp. 939-952, 2013.
- [27] S. J. Bae, T. Yuan, S. Ning, and W. Kuo, "A Bayesian approach to modeling two-phase degradation using change-point regression," *Reliability Engineering & System Safety*, vol. 134, pp. 66-74, 2015.
- [28] J. Feng, Q. Sun, and T. Jin, "Storage life prediction for a high-performance capacitor using multi-phase Wiener degradation model," *Communications in Statistics-Simulation and Computation*, vol. 41, pp. 1317-1335, 2012.
- [29] X. Wang, P. Jiang, B. Guo, and Z. Cheng, "Real-time Reliability Evaluation for an Individual Product Based on Change-point Gamma and Wiener Process," *Quality and Reliability Engineering International*, vol. 30, pp. 513-525, 2014.
- [30] Y. Wen, J. Wu, and Y. Yuan, "Multiple-Phase Modeling of Degradation Signal for Condition Monitoring and Remaining Useful Life Prediction," *IEEE Transactions on Reliability*, vol. 66, pp. 924-938, 2017.

- [31] T. S. Ng, "An application of the EM algorithm to degradation modeling," *IEEE Transactions on Reliability*, vol. 57, pp. 2-13, 2008.
- [32] D. Kong, N. Balakrishnan, and L. Cui, "Two-Phase Degradation Process Model With Abrupt Jump at Change Point Governed by Wiener Process," *IEEE Transactions on Reliability*, 2017.
- [33] W.-a. Yan, B.-w. Song, G.-l. Duan, and Y.-m. Shi, "Real-time reliability evaluation of two-phase Wiener degradation process," *Communications in Statistics-Theory and Methods*, vol. 46, pp. 176-188, 2017.
- [34] Z.-S. Ye, Y. Wang, K.-L. Tsui, and M. Pecht, "Degradation data analysis using Wiener processes with measurement errors," *IEEE Transactions on Reliability*, vol. 62, pp. 772-780, 2013.
- [35] P. Fearnhead and Z. Liu, "Efficient Bayesian analysis of multiple changepoint models with dependence across segments," *Statistics and Computing*, vol. 21, pp. 217-229, 2011.
- [36] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Selected Papers of Hirotugu Akaike*, ed: Springer, 1998, pp. 199-213.
- [37] G. Schwarz, "Estimating the dimension of a model," *The annals of statistics*, vol. 6, pp. 461-464, 1978.
- [38] A. Hannart and P. Naveau, "An improved Bayesian information criterion for multiple change-point models," *Technometrics*, vol. 54, pp. 256-268, 2012.
- [39] X.-S. Si, W. Wang, C.-H. Hu, M.-Y. Chen, and D.-H. Zhou, "A Wiener-process-based degradation model with a recursive filter algorithm for remaining useful life estimation," *Mechanical systems and signal processing*, vol. 35, pp. 219-237, 2013.