Multiple-Change-Point Modeling and Exact Bayesian Inference of Degradation Signal for Prognostic Improvement

Yuxin Wen

Jianguo Wu

Qiang Matthew Zhou

Tzu-Liang Bill Tseng

Multiple-Change-Point Modeling and Exact Bayesian Inference of Degradation Signal for Prognostic Improvement

Comments

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in *IEEE Transactions on Automation Science and Engineering*, volume 16, issue 2, in 2018 following peer review. The definitive publisher-authenticated version is available online at https://doi.org/10.1109/TASE.2018.2844204.

Copyright

Multiple Change-point Modeling and Exact Bayesian Inference of Degradation Signal for Prognostics Improvement

Yuxin Wen, Jianguo Wu[1], Qiang Zhou, Tzu-Liang (Bill) Tseng

Abstract— Prognostics plays an increasingly important role in modern engineering systems for smart maintenance decision-making. In the parametric regression-based approach, the parametric model is often too rigid to model degradation signals in many applications. In this paper, we propose a Bayesian multiple change-point modeling framework to better capture the degradation path and improve the prognostics. At the offline modeling stage, a novel stochastic process is proposed to model the joint prior of change-points and positions. All hyperparameters are estimated through an empirical two-stage process. At the online monitoring and remaining useful life (RUL) prediction stage, a recursive updating algorithm is developed to exactly calculate the posterior distribution and RUL prediction sequentially. To control the computational cost, a fixed-support-size strategy in the online model updating and a partial Monte Carlo strategy in the RUL prediction are proposed. The effectiveness and advantages of the proposed method are demonstrated through thorough simulation and real case studies.

Note to Practitioners— Degradation signals have been widely used in determining the current health condition and estimate the remaining useful life (RUL) of a component or a system. Most of the existing prognostics utilize a parametric regression model to describe the evolution path of degradation signals for RUL prediction. The common functional forms of these models include simple linear, quadratic, and exponential functions. However, in many applications, the degradations signals show multiple segments characteristics and the existing parametric forms are inadequate to capture the degradation trend. Motivated by such issue, this paper presents a multiple change-point modeling approach, where the degradation signal is divided into several consecutive segments by change-points, and each segment is modeled by a unique parametric model. To capture the heterogeneity across different units, all the parameters, including the number and locations of change-points, model parameters of each segment, are assumed to be random variables following certain distributions. Then we develop a statistical method to estimate these distributions using historical data. At the online monitoring stage, we develop an innovative updating algorithm to exactly calculate the closed forms of the posterior distributions of the latest change-point, the current segment, and model parameters of the current segment. We also derive a closed form for the RUL distribution estimation. Later several efficient approximation strategies are proposed to reduce the computational burden. Simulation studies and real case studies have shown that the proposed methodology has much better performance than existing approaches in handling degradation signals of multiple-segment characteristics. In future research, we will extend the multiple change-point modeling approach to stochastic process based prognostics, such as Wiener process.

Index Terms—Multiple change-point model, exact Bayesian inference, remaining useful life prediction, prognostics, degradation modeling.

I. INTRODUCTION

PROGNOSTICS refers to the process of evaluating the current health of a system or a subcomponent and then predicting the remaining useful life (RUL) based on the current health condition [1]. It has played an increasingly important role in modern engineering systems and manufacturing processes due to its capability of reducing maintenance costs, improving operational efficiency and facilitating decision-making [2]. The prediction of RUL often requires a prognostic model, which can be generally classified into two groups, physical-based and data-driven based models [3]. The physical models require a complete understanding of the specific degradation mechanisms and are often infeasible or ineffective in practical applications due to high system complexity or unclear degrading mechanism [4]. On the other hand, the data-driven approaches often make use of condition monitoring (CM) data for prognostics and become more and more popular. CM signals, also known as degradation signals, are closely related with the underlying system degradation processes. They are collected in-situ through sensors during the system operations and provide great opportunities to monitor the health condition and predict the future failures. A common assumption in the CM signal based prognostics is that the system or subcomponent is considered failed once the CM signal crosses

[1] Corresponding author: Jianguo Wu

Y. Wen is with the Department of Electrical and Computer Engineering (ECE), University of Texas at El Paso, Texas 79968, USA (e-mail: ywen@miners.utep.edu).

J. Wu is with the Department of Industrial, Manufacturing and Systems Engineering (IMSE), and ECE Department (affiliated), University of Texas at El Paso, Texas 79968, USA (e-mail: jwu2@utep.edu).

Q. Zhou is with the Department of Systems and Industrial Engineering, The University of Arizona, Tucson, AZ 85721, USA (e-mail: zhouq@email.arizona.edu).

B. Tseng is with the IMSE Department, University of Texas at El Paso, Texas 79968, USA (e-mail: btseng@utep.edu).

a predefined failure threshold. With the fast development of sensing and condition monitoring technologies, there has been a rapid increase in the research of CM signal based prognostics over the past few years [2]. The existing data-driven prognostics that utilize CM signals can be grouped into two categories [3, 5, 6]: artificial intelligence (AI) techniques and statistical approaches. The typical AI techniques include the neural networks, support vector machines (SVM), decision tree, and fuzzy logic system, while the statistical approaches include various stochastic processes (e.g., Wiener process, Gamma process and inverse Gaussian process), state space models, and regression-based models, etc. The statistical approaches are more commonly used than AI techniques due to their excellent statistical properties in various aspects, e.g., interpretation and uncertainty quantification. A comprehensive review of the state-of-the-art in statistical methods for prognostics is given by Si *et al* [7] .

The regression-based approach is one of the most natural ways to capture the evolution path of the degradation signals [8-10]. In these models, a parametric function, such as linear, polynomial and exponential function, is applied to model the population trend. The regression coefficients are often assumed to be random in order to characterize the individual heterogeneity. Unfortunately, these functional forms are often too simple or inadequate to model the whole degradation signals in many practical applications. Indeed, the real degradation signals are often irregular and show multiple phases where each phase needs a regression function. For example, Son *et al.* [11] showed that the resistance of vehicle batteries have obvious two phases where each phase can be well modelled by a quadratic function. Another well-known example is the bearing vibrational signal [12, 13], where a stable stage and a rapidly increasing stage can easily be observed. Bae and Kvam [14] demonstrated that the degradation path of vacuum fluorescent displays is not monotonic and it contains two or even three phases. Other examples include the degradation data for semiconductor laser diodes [15], high-performance capacitors [16], and the liquid coupling devices [17].

To deal with this problem, various efforts have been made. Some researchers proposed to ignore the observations in the first phase, and then build a parametric model only based on the remaining data [8, 18, 19]. The RUL prediction is performed only at the second stage, based on the assumption that failures will not occur at the early stage. However, this strategy has two severe drawbacks. Firstly, the truncated data may contain valuable information about the future degradation path, and thus should be utilized to improve the prognostic accuracy. Secondly, the change-points that separate the two phases are treated as deterministic in these methods. However, they are random and vary across different units in practice. To fully utilize the observations in the early stage, some researchers proposed a two-phase model with a random change-point [11, 13, 20, 21]. For instance, Son *et al.* [11] incorporated a change-point to the resistance signal in a joint prognostic model (joint modeling of reliability data and CM data) to predict the RUL of batteries. In their method, the concordance correlation

coefficient (CCC) criterion is employed for online detection of the change-point. They found that the addition of a change-point improved the accuracy of prediction. Chen and Tsui [13] developed a two-phase model by extending Gebraeel *et al.*'s work [8], and a two-step empirical method is used for online change-point detection. However, although those models with one change-point improved the prediction accuracy for some specific cases, they may be inadequate or inapplicable to more complex degradation signals with three or even more degradation phases. Besides, the online detection of the change-point is often based on some heuristics, and is not fully incorporated into the Bayesian framework. To the best of our knowledge, very limited work with multiple phases is available in the existing literature, especially in the regression-based approaches. Feng *et al.* [16] proposed a multi-phase Wiener process model to predict the storage life of high-voltage-pulse capacitors. However, in their work, the number of change-points and their locations are deterministic and the same for all units, which is not realistic for real degradation signals with unit heterogeneity.

Motivated by the aforementioned issues, this paper proposes a novel multiple change-point modelling approach to better capture the degradation path for prognostics improvement. To characterize the inherent unit-to-unit heterogeneity and make the model more flexible, all the model parameters are assumed to be random, including the number of change-points, their locations, and all model parameters of each segment. In the regression-based prognostics, two stages are often required: the off-line modeling of historical CM data, and the online Bayesian individual model updating and RUL prediction of a new unit. The estimated parameters of these random distributions in the off-line stage are used as hyperparameters in the second stage, where both the information of the historical data and the current health condition are combined for prognostics. However, due to the greatly increased model dimensionality and complexity, there are several critical challenges to be addressed. The first challenge is to specify appropriate priors for the change-points and how to estimate the hyperparameters efficiently. To address this issue, a novel stochastic process is proposed to model the occurrence of change-points and then a series of approaches are developed to estimate the hyperparameters in this paper. The second challenge lies in the online stage, where the posterior distributions of the total number of change-points on the CM signal, the number of change-points occurred, the location of the latest change-point, and the model parameters of the current segment have to be sequentially calculated for RUL prediction. The particle filters (PFs) [22-24] are currently standard ways in sequential Bayesian inference of multiple change-point models, due to their capability of tracking highly nonlinear state-space models. However, the PFs are notorious for their significant complexity and computational cost in handling high dimensional problems, which may limit their applications in online monitoring and prognostics. In this paper, we develop an innovative recursive updating algorithm to overcome this challenge, where the exact Bayesian inference or the closed form of all posterior distributions can be sequentially obtained.

The third major challenge is the RUL prediction. Due to the unknown number and locations of future change-points, the exact calculation of the RUL distribution is very complex and time-consuming. We derive a closed form of the RUL prediction and then propose a partial Monte Carlo (partial-MC) approach to control the computational cost of RUL estimation.

The remainder of the paper is organized as follows. In Section 2, a multiple change-point model for the CM signals is presented. The prior specification and parameter estimation for the multiple change-point model is given in Section 3. Section 4 presents the technical details on how to sequentially update the posterior distributions of all necessary model parameters and how to perform RUL prediction. Section 5 demonstrates the effectiveness and accuracy of the proposed method through numerical and case studies. The conclusion and discussion are given in Section 6.

II. MULTIPLE CHANGE-POINT MODELING OF DEGRADATION SIGNAL

A degradation model that can adequately describe the degradation path is essential for prognostics. In this section, we give the details of the multiple change-point model used in this paper. Specifically, we model the degradation signals with piecewise linear regression models, e.g., each segment is modelled as a polynomial of a certain order. With a proper number of change-points at certain locations, this model is capable of capturing both the non-linear and multiple-phase characteristics of various degradation signals. For illustration, we use thrust bearing vibration signals [8] with multiple-phase characteristic, as shown in Fig. 1. Obviously, the bearing operates under a stable condition at first and then degrades rapidly with two distinct phases. The degradation signal could be appropriately modelled with three line segments. If only one change-point is incorporated, the degradation signal after the stable stage is poorly fitted (Fig. 1a), which could consequently influence the prognostic accuracy (Fig. 1c).
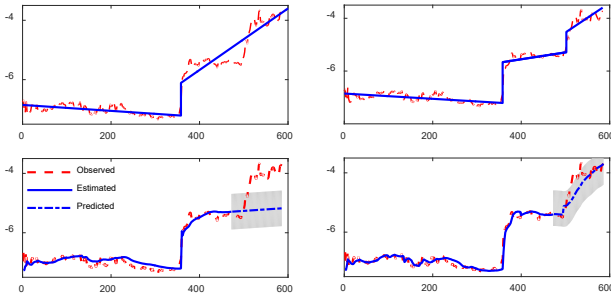

Fig. 1. Modeling and prediction of degradation signal with two line segments (a and c) and three line segments (b and d). The dark regions are prediction confidence intervals.

Suppose there are I historical CM signals. Let $\boldsymbol{Y}_i$ denote the CM signal of the i-th unit, and $\boldsymbol{Y}_i = \{y_{i,1}, y_{i,2}, \ldots, y_{i,j}, \ldots, y_{i,n_i}\}, i = 1, \ldots, I$ where $y_{i,j}$ is the j-th observation of unit i at time $t_{i,j}$, and n_i is the total number of observations in the lifetime. Let k_i denote the total number of change-points of unit i before failure, which is modelled as a

random variable to account for the unit-to-unit heterogeneity. Following a conventional notation of multiple change-point models [25], suppose the k_i change-points are the integer-valued indices $c_{i1}, c_{i2}, \ldots, c_{ik_i}$. For notational convenience, we define $c_{i0} = 0$ and $c_{ik_i+1} = n_i$. Then $c_{i0} = 0 < c_{i1} < c_{i2} < \cdots < c_{ik_i} < c_{ik_i+1} = n_i$. Consequently, the sequence of observations $\{y_{i,1}, y_{i,2}, \ldots, y_{i,j}, \ldots, y_{i,n_i}\}$ are partitioned into $k_i + 1$ contiguous segments $y_{i,c_{i0}+1:c_{i1}}, y_{i,c_{i1}+1:c_{i2}}, \ldots y_{i,c_{ik_i}+1:n_i}$. Mathematically, the multiple change-point model can be expressed as

$$y_{i,j} = \begin{cases} \boldsymbol{X}_{i,j,1}\boldsymbol{\beta}_i^{(1)} + \sigma_i^{(1)}\varepsilon_{ij}, & if\ 0 < t_{i,j} \le t_{i,c_{i1}} \\ \boldsymbol{X}_{i,j,2}\boldsymbol{\beta}_i^{(2)} + \sigma_i^{(2)}\varepsilon_{ij}, & if\ t_{i,c_{i1}} < t_{i,j} \le t_{i,c_{i2}} \\ \quad\quad\cdots \\ \boldsymbol{X}_{i,j,k_i}\boldsymbol{\beta}_i^{(k_i)} + \sigma_i^{(k_i)}\varepsilon_{ij}, & if\ t_{i,c_{ik_i-1}} < t_{i,j} \le t_{i,c_{ik_i}} \\ \boldsymbol{X}_{i,j,k_i+1}\boldsymbol{\beta}_i^{(k_i+1)} + \sigma_i^{(k_i+1)}\varepsilon_{ij}, & if\ t_{i,c_{ik_i}} < t_{i,j} \le t_{i,n_i} \end{cases} \quad (1)$$

where $\boldsymbol{X}_{i,j,s}$ is the vector of polynomial basis functions, i.e., $\boldsymbol{X}_{i,j,s} = \left[1, (t_{i,j} - t_{i,c_{is-1}}), \cdots, (t_{i,j} - t_{i,c_{is-1}})^{q_{is}}\right]$ where q_{is} is the polynomial order of the s-th segment, $\boldsymbol{\beta}_i^{(s)}$ is a vector of regression parameter and $\sigma_i^{2(s)}$ is noise variance of the s-th segment, and ε_{ij} is a noise term following *i.i.d.* standard normal distribution. Note that the order of polynomial regression could vary across different segments. Give the position of a change-point, we assume that the observations before that change-point is independent of those after the change-point. For simplicity and without loss of generality, we assume that $t_{i,j} = j$ in the rest of the paper, i.e., the sampling intervals equal to 1 for all units. Besides, given the total number of change-points k_i, we assume that the polynomial orders of the $k_i + 1$ segments are deterministic.

The prognostics often involves two stages, namely, the offline stage for modeling and estimation, and the online stage for sequential model updating and RUL prediction. To characterize both the population trend and the individual heterogeneity, all the model parameters are assumed random in the offline modeling of the historical data. Denote a multiple change-point model as $\boldsymbol{\mathcal{M}} = \left(k, \{\delta^{(s)}\}_{s=1}^{k+1}, \{\boldsymbol{\theta}^{(s)}\}_{s=1}^{k+1}\right)$ where k is the number of change-points, $\delta^{(s)} = c_s - c_{s-1}$ is the duration of the sth segment, and $\boldsymbol{\theta}^{(s)} = \left(\boldsymbol{\beta}^{(s)}, \sigma^{2(s)}\right)$ is the model parameters of the sth segment. In the offline modeling, all these parameters are modelled with appropriate distributions, and the hyperparameters are estimated. The estimated distributions are then used as priors in the online Bayesian model updating and RUL prediction. At the online stage, the posterior distributions of the individual model parameters are sequentially updated. The total number of change-points, the index of the current segment (or how many change-points have occurred), and the latest change-point (or the duration since the latest change-point) are three key parameters in Bayesian model updating and RUL prediction.

Let $y_{1:t}$ denote the observations of a working unit up to the

current time t, and $\boldsymbol{x}_t = (\boldsymbol{\theta}_t, \tau_t, s_t, k)$ be the state vector where $\boldsymbol{\theta}_t = (\boldsymbol{\beta}_t, \sigma_t^2)$ are the model parameters of the current segment, τ_t be the latest change-point that has occurred ($\tau_t \leq t - 1$), and s_t be index of the current segment, e.g., $s_t = 1, 2, \dots, k + 1$. At the online stage, the posterior distribution $p(\boldsymbol{x}_t | y_{1:t})$ and the predictive density $p(y_{t+L} | y_{1:t})$ for integer $L > 0$ have to be calculated. However, these distributions are generally intractable. As mentioned earlier, although the particle filtering techniques or sequential Monte Carlo techniques are capable of handling these nonlinear intractable problems, their efficiency is significantly affected by their notorious particle degeneracy and impoverishment issues, and the enormous computational cost [22]. In this paper, we find that by using conjugate priors, the closed form of the posterior distributions and the predictive density can be recursively and efficiently calculated. The overall prognostic framework with offline and online stages is summarized in Fig. 2. The following two sections give the technical details about the prior specification and parameter estimation at the offline stage, and the posterior distribution estimation and RUL prediction at the online stage.
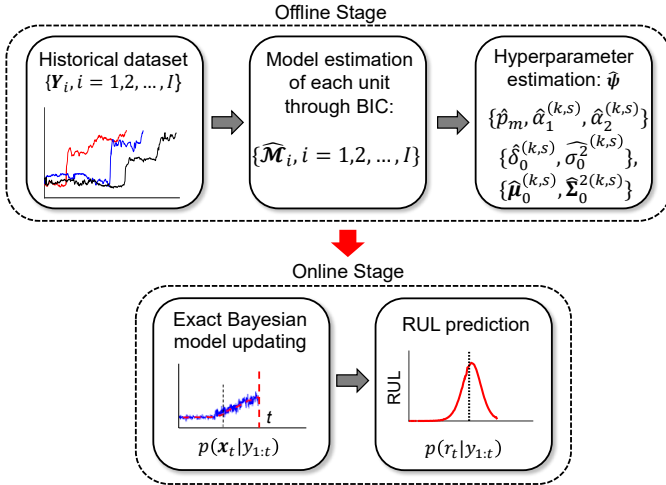

Fig. 2. Illustration of the proposed prognostic framework.

III. Prior Specification and Parameter Estimation

A. Specification of Priors

Prior distribution plays an important role in Bayesian data analysis. Informative priors are often preferred if historical data is available, since they reflect the strong belief of a new unit and can lead to more accurate posterior inference of the degradation path. In this section, the priors for the multiple change-point model $\boldsymbol{\mathcal{M}} = \left(k, \{\delta^{(s)}\}_{s=1}^{k+1}, \{\boldsymbol{\theta}^{(s)}\}_{s=1}^{k+1}\right)$ are specified and estimated.

In the existing literature, the change-points in the multiple change-point models are often modelled through a Markov process where the occurrence of the next change-point only depends on the duration since the occurrence of the previous one [22, 26-29]. This process indirectly specifies a joint prior on the number of change-points and the durations between successive change-points for a given time series data. For example, when the durations are continuous variables, a Poisson process can be applied, where the durations $\delta^{(s)}, s =$

$1, \dots, k + 1$ follow an *i.i.d.* exponential distribution [28]. The joint prior of k and $\{\delta^{(s)}\}_{s=1}^{k+1}$ for a time series data of duration T can be easily derived as

$$\pi\left(k, \{\delta^{(s)}\}_{s=1}^{k+1}\right) = \left[\prod_{s=1}^{k} \pi(\delta^{(s)} | \lambda)\right] P(\delta^{(k+1)} \geq T - c_k) \tag{2}$$
$$= \lambda^k \exp(-\lambda T)$$

where λ is the Poisson rate. If the durations $\delta^{(s)}$ are positive integers or the change-point locations are the observation indices, which is most conventional in the existing literature, a Bernoulli process is often applied, or equivalently a geometric distribution is applied to the durations [22, 26, 27, 29]. The joint density is simply $\pi\left(k, \{\delta^{(s)}\}_{s=1}^{k+1}\right) = p^k (1 - p)^{n-1-k}$ where p is the parameter of the Bernoulli distribution and n is the total number of observations. However, such homogeneous Markov process does not fit the degradation signals well. Firstly, it does not consider the segment heterogeneity. To make the prior more informative, each segment should has its own distribution parameters for the duration. Secondly, the support size of the discrete variable k increases with the total number of observations, which is unrealistic for degradation modeling, where the maximum number of segments should be controlled. In this paper, we propose a special nonhomogeneous Markov process where discrete distributions that are independent of the number of observations are selected for the number of change-points k, and then the positions of the change-points are modelled as a nonhomogeneous Markov process with durations between successive change-points depending on both k and the segment index s. The joint prior for both change-points and model parameters could be formulated as

$$\pi(\boldsymbol{\mathcal{M}}) = \pi(k) \prod_{s=1}^{k} \pi(\delta^{(s)} | k) \prod_{s=1}^{k+1} \pi(\boldsymbol{\theta}^{(s)} | k) \tag{3}$$

In this paper, we select a categorical distribution for k which is independent of n, i.e., $\pi(k = m) = p_m$ and $\sum_m p_m = 1$. The phase durations, which are discrete in this paper, are approximately modelled with continuous distributions, as they are more flexible in controlling mean and variance than most of the existing discrete probability distributions. ==For the sake of simplicity and without loss of generality, we assume that the duration of each phase follows a normal distribution,== $\delta^{(s)} | k \sim N(\delta_0^{(k,s)}, \sigma_0^{2(k,s)})$. For the model parameters of each phase, the commonly used normal and inverse Gamma (IG) conjugate priors are assumed:

$$\pi(\boldsymbol{\beta}^{(s)}, \sigma^{2(s)} | k) = \pi(\sigma^{2(s)} | k) \pi(\boldsymbol{\beta}^{(s)} | \sigma^{2(s)}, k)$$
$$= IG(\sigma^{2(s)} | \alpha_1^{(k,s)}, \alpha_2^{(k,s)}) N(\boldsymbol{\beta}^{(s)} | \boldsymbol{\mu}_0^{(k,s)}, \sigma^{2(s)} \boldsymbol{\Sigma}_0^{(k,s)}) \tag{4}$$

For notational convenience, in the rest of the paper we use the double superscript (k, s) to denote the parameter or variable of

s th segment conditioning that there are in total k change-points, e.g., $\boldsymbol{\beta}^{(k,s)} = \boldsymbol{\beta}^{(s)}|k$.

B. Parameter Estimation from Historical Data

Informative priors can be obtained by estimating all the hyperparameters through historical data. Let $\boldsymbol{\psi}$ denote the vector of all hyperparameters, i.e., $\{p_m\}, \left\{\alpha_1^{(k,s)}, \alpha_2^{(k,s)}\right\}$ and $\{\boldsymbol{\mu}_0^{(k,s)}, \boldsymbol{\Sigma}_0^{(k,s)}\}$. A natural way to estimate $\boldsymbol{\psi}$ is to maximize the marginal likelihood of I historical CM signals [13]:

$$\widehat{\boldsymbol{\psi}} = \arg\max_{\boldsymbol{\psi}} \prod_{i=1}^{I} \int p(\boldsymbol{Y}_i|\boldsymbol{\mathcal{M}}_i)\pi(\boldsymbol{\mathcal{M}}_i|\boldsymbol{\psi})\, d\boldsymbol{\mathcal{M}}_i \tag{5}$$

However, the marginal likelihood above is very complex and generally intractable. ==There are two approaches to address this issue, one being the expectation-maximization (EM) algorithm [15], where the missing variables are the change-point locations and the changing model parameters of each degradation signal, and the other one being the empirical two-stage estimation method, which is a simpler alternative to EM algorithm for the random effects models or the Bayesian hierarchical models [11, 13, 30]. Compared with the EM algorithm, the empirical two-stage estimation is much easier and more efficient to implement. Although some biases may be introduced [31], it is often negligible according to the comparison study [32]. Therefore, for simplicity, we use the empirical two-stage estimation method.== Specifically, the model parameters $\widehat{\boldsymbol{\mathcal{M}}}_i$ of each historical unit $i = 1, \dots, I$ are obtained at the first stage, and then the estimated parameters $\{\widehat{\boldsymbol{\mathcal{M}}}_i, i = 1, \dots, I\}$ are treated as observed values in the estimation of hyperparameters through the maximum likelihood estimation (MLE) approach.

For multiple change-point models with unknown change-points, the typical MLE is not applicable for parameter estimation, since increasing the number of change-points will always increase the likelihood and result in over-fitting issue. To address this issue, we use the Bayesian information criterion (BIC) [33] for change-point model selection and the corresponding segment parameter estimation. For notational convenience, we ignore the subscript i for individual unit in the following four equations. The parameter estimation for each unit can be formulated as

$$\widehat{\boldsymbol{\mathcal{M}}} = \arg\min_{\boldsymbol{\mathcal{M}}} (-2l(\boldsymbol{\mathcal{M}}|\boldsymbol{Y}) + K\log n) \tag{6}$$

where K is the total number of parameters, including change-points, regression parameters, and noise variances, and $l(\boldsymbol{\mathcal{M}}|\boldsymbol{Y})$ is the log-likelihood function expressed by

$$l(\boldsymbol{\mathcal{M}}|\boldsymbol{Y}) = \sum_{s=1}^{k+1} \left[-\frac{1}{2}(c_s - c_{s-1})\log(2\pi\sigma^{2(k,s)}) \right. $$
$$\left. - \frac{\left\| y_{c_{s-1}+1:c_s}^T - \boldsymbol{X}_{1,c_s-c_{s-1}} \boldsymbol{\beta}^{(k,s)} \right\|^2}{2\sigma^{2(k,s)}} \right] \tag{7}$$

where $\boldsymbol{X}_{1,c_s-c_{s-1}}$ is the design matrix, with $\boldsymbol{X}_{t_1,t_2}$ of order q defined as

$$\boldsymbol{X}_{t_1,t_2} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_1+1 & \cdots & t_2 \\ \cdots & \cdots & \cdots & \cdots \\ t_1^q & (t_1+1)^q & \cdots & t_2^q \end{bmatrix}^T \tag{8}$$

Conditioning on the fixed change-points, i.e., $\left(k, \{\delta^{(k,s)}\}_{s=1}^{k+1}\right)$, the parameters of each segment that minimize Eq. (6) can be easily obtained through MLE of classical linear models

$$\begin{aligned} \widehat{\boldsymbol{\beta}}^{(k,s)} &= \left(\boldsymbol{X}_{1,c_s-c_{s-1}}^T \boldsymbol{X}_{1,c_s-c_{s-1}}\right)^{-1} \boldsymbol{X}_{1,c_s-c_{s-1}}^T y_{c_{s-1}+1:c_s}^T, \\ \widehat{\sigma^2}^{(k,s)} &= \left\| y_{c_{s-1}+1:c_s}^T - \boldsymbol{X}_{1,c_s-c_{s-1}} \widehat{\boldsymbol{\beta}}^{(k,s)} \right\|^2 / \delta^{(k,s)} \end{aligned} \tag{9}$$

Therefore, for each possible model defined by change-points, the BIC value can be easily evaluated. Suppose the estimated parameters are $\widehat{\boldsymbol{\mathcal{M}}}_i = \left(\widehat{k}_i, \left\{\widehat{\delta}_i^{(\widehat{k}_i,s)}\right\}_{s=1}^{\widehat{k}_i}, \left\{\widehat{\boldsymbol{\beta}}_i^{(\widehat{k}_i,s)}, \widehat{\sigma^2}_i^{(\widehat{k}_i,s)}\right\}_{s=1}^{\widehat{k}_i+1}\right)$ for $i = 1, \dots, I$. The second stage is to estimate the hyperparameters based on $\widehat{\boldsymbol{\mathcal{M}}}_i, i = 1, \dots, I$ through MLE approach. The MLE of hyperparameters p_m can be easily obtained as

$$\widehat{p}_m = \frac{1}{I} \sum_{i=1}^{I} \mathbf{1}_{\widehat{k}_i=m} \tag{10}$$

For the hyperparameters $(\alpha_1^{(k,s)}, \alpha_2^{(k,s)})$ in the inverse Gamma distribution, the MLE using observations $\left\{\widehat{\sigma^2}_i^{(\widehat{k}_i,s)} \middle| \widehat{k}_i = k, i = 1, \dots, I\right\}$ can be estimated numerically through various optimization algorithms. The MLE of the hyperparameters $(\delta_0^{(k,s)}, \sigma_0^{2(k,s)})$ for the segment duration can be obtained straightforwardly as

$$\begin{aligned} \widehat{\delta}_0^{(k,s)} &= \frac{\sum_{i=1}^{I} \delta_i^{(\widehat{k}_i,s)} \mathbf{1}_{\widehat{k}_i=k} / \sum_{i=1}^{I} \mathbf{1}_{\widehat{k}_i=k}}{\sum_{i=1}^{I} \mathbf{1}_{\widehat{k}_i=k}} \\ \widehat{\sigma_0^2}^{(k,s)} &= \frac{\sum_{i=1}^{I} \left(\delta_i^{(\widehat{k}_i,s)} - \delta_0^{(k,s)}\right)^2 \mathbf{1}_{\widehat{k}_i=k}}{\sum_{i=1}^{I} \mathbf{1}_{\widehat{k}_i=k}} \end{aligned} \tag{11}$$

For the hyperparameters $\boldsymbol{\mu}_0^{(k,s)}$ and $\boldsymbol{\Sigma}_0^{(k,s)}$, closed forms can be derived as

$$\widehat{\boldsymbol{\mu}}_0^{(k,s)} = \sum_{i=1}^{I} \frac{\widehat{\boldsymbol{\beta}}_i^{(\widehat{k}_i,s)} \mathbf{1}_{\widehat{k}_i=k}}{\widehat{\sigma^2}_i^{(\widehat{k}_i,s)}} \bigg/ \sum_{i=1}^{I} \frac{\mathbf{1}_{\widehat{k}_i=k}}{\widehat{\sigma^2}_i^{(\widehat{k}_i,s)}}$$

$$\widehat{\boldsymbol{\Sigma}}_0^{(k,s)} = \frac{\sum_{i=1}^{I} \left(\widehat{\boldsymbol{\beta}}_i^{(\widehat{k}_i,s)} - \widehat{\boldsymbol{\mu}}_0^{(k,s)}\right)\left(\widehat{\boldsymbol{\beta}}_i^{(\widehat{k}_i,s)} - \widehat{\boldsymbol{\mu}}_0^{(k,s)}\right)^T \mathbf{1}_{\widehat{k}_i=k}}{\sum_{i=1}^{I} \mathbf{1}_{\widehat{k}_i=k}} \tag{12}$$

The details of the derivation can be found in Appendix A.

IV. Exact Bayesian Online Model Updating and RUL Prediction

Once the prior information from the historical data is obtained, the model updating (i.e., posterior inference) and RUL prediction can be conducted for a specific in-service unit at the online stage. In this section we will discuss how to update the model sequentially through exact Bayesian inference and how to predict the RUL for a new in-service unit.

A. Exact Bayesian Online Model Updating

The model updating of a working unit is an essential step for health condition monitoring and RUL prediction. It refers to the posterior distribution evaluation of all model parameters that could capture the current health condition and future degradation evolution. In this paper, the model updating is to calculate the posterior distribution $p(x_t|y_{1:t})$ where the state vector $x_t = (\theta_t, \tau_t, s_t, k)$. Generally, this posterior distribution is intractable and sequential Monte Carlo techniques are needed. Fortunately, due to the assignment of conjugate priors for θ, the posterior could be recursively calculated, which is shown as follows.

To calculate the joint posterior $p(x_t|y_{1:t})$, we first calculate the posterior distribution of the discrete components $P(\tau_t, s_t, k|y_{1:t})$, and then calculate the posterior distribution of the continuous components conditioning on the discrete components, i.e., $p(\theta_t|\tau_t, s_t, k, y_{1:t})$. That is

$$p(x_t|y_{1:t}) = P(\tau_t, s_t, k|y_{1:t})p(\theta_t|\tau_t, s_t, k, y_{1:t}) \quad (13)$$

The conditional posterior distribution of the continuous components $p(\theta_t|\tau_t, s_t, k, y_{1:t})$ can be calculated based on Theorem 1 as follows.

Theorem 1 Suppose the conjugate prior in Eq.(4) is assigned to β_t and σ_t^2.

$$\pi(\beta_t, \sigma_t^2|s_t = s, k) = IG\left(\sigma_t^2\middle|\alpha_1^{(k,s)}, \alpha_2^{(k,s)}\right) N\left(\beta_t\middle|\mu_0^{(k,s)}, \sigma_t^2\Sigma_0^{(k,s)}\right)$$

Then

$$(\sigma_t^2|\tau_t = j, s_t = s, k, y_{1:t}) \sim IG\left(\alpha_1^{(k,s)} + \frac{t-j}{2}, \alpha_2^{(k,s)} + \frac{H_{j+1,t}^{(k,s)}}{2}\right)$$

$$(\beta_t|\sigma_t^2, \tau_t = j, s_t = s, k, y_{1:t}) \sim N(\mu_{j+1,t}, \sigma_t^2\Sigma_{j+1,t}) \quad (14)$$

where

$$\Sigma_{j+1,t} = \left(X_{1,t-j}^T X_{1,t-j} + \left(\Sigma_0^{(k,s)}\right)^{-1}\right)^{-1}$$
$$N_{j+1,t} = \left(\left(\Sigma_0^{(k,s)}\right)^{-1}\mu_0^{(k,s)} + X_{1,t-j}^T y_{j+1:t}\right)$$
$$\mu_{j+1,t} = \Sigma_{j+1,t} N_{j+1,t} \quad (15)$$
$$H_{j+1,t}^{(k,s)} = y_{j+1:t}^T y_{j+1:t} + \left(\mu_0^{(k,s)}\right)^T \left(\Sigma_0^{(k,s)}\right)^{-1}\mu_0^{(k,s)}$$
$$- N_{j+1,t}^T \Sigma_{j+1,t} N_{j+1,t}$$

The proof of Theorem 1 can be found in Appendix B. The calculation of $P(\tau_t, s_t, k|y_{1:t})$ is the main challenge. It can be recursively updated as

$$P(\tau_t = j, s_t = s, k|y_{1:t})$$
$$\propto P(\tau_t = j, s_t = s, k|y_{1:t-1})p(y_t|\tau_t = j, s_t = s, k, y_{1:t-1}) \quad (16)$$

Eq. (16) consists of two parts. The first part is the predictive probability mass function (PMF) $P(\tau_t = j, s_t = s, k|y_{1:t-1})$ which can be recursively calculated by

$$P(\tau_t = j, s_t = s, k|y_{1:t-1})$$
$$= \sum_{j',s'} P(\tau_{t-1} = j', s_{t-1} = s', k|y_{1:t-1})P(\tau_t = j, s_t \quad (17)$$
$$= s|\tau_{t-1} = j', s_{t-1} = s', k, y_{1:t-1})$$

where $P(\tau_{t-1} = j', s_{t-1} = s', k|y_{1:t-1})$ is the posterior distribution obtained at the previous time step, and $P(\tau_t = j, s_t = s|\tau_{t-1} = j', s_{t-1} = s', k, y_{1:t-1})$ is the predictive Markov transition probability. Based on the specified nonhomogeneous Markov process for change-points, this predictive Markov transition probability can be derived as

$$P(\tau_t = j, s_t = s|\tau_{t-1} = j', s_{t-1} = s', k, y_{1:t-1})$$
$$= \begin{cases} \dfrac{1 - G^{(k,s')}(t-j')}{1 - G^{(k,s')}(t-j'-1)}, & \text{if } j = j' \text{ and } s = s' < k+1 \\ 1, & \text{if } j = j' \text{ and } s = s' = k+1 \\ \dfrac{G^{(k,s')}(t-j') - G^{(k,s')}(t-j'-1)}{1 - G^{(k,s')}(t-j'-1)}, & \text{if } j = t-1 \text{ and } s = s'+1 \\ & \qquad\qquad \leq k+1 \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

where $G^{(k,s')}(\cdot)$ is the cumulative distribution function of s'-th segment duration for a CM signal with k change-points.

The second part of Eq. (16) is $p(y_t|\tau_t = j, s_t = s, k = m, y_{1:t-1})$, namely, the predictive density function of y_t. It is the only term that involves the newest observation y_t in the posterior updating of the discrete components. Once this density function is known, the posterior distribution of the discrete components can be recursively updated based on Eq. (16), (17) and (18). Therefore this part is critically important. It can be calculated based on Theorem 2 as follows.

Theorem 2 Denote $X_t = [1, t, \cdots, t^q]$ where q is the polynomial order, then if $j < t-1$,

$$(y_t|\tau_t = j, s_t = s, k, y_{j+1:t-1}) \sim$$
$$t_1\left(2\alpha_1^{(k,s)} + t - j - 1, X_{t-j}\mu_{j+1,t-1}, \frac{2\alpha_2^{(k,s)} + H_{j+1:t-1}^{(k,s)}}{2\alpha_1^{(k,s)} + t - j - 1}\left(1 \right.\right.$$
$$\left.\left. + X_{t-j}\Sigma_{j+1,t-1}X_{t-j}^T\right)\right) \quad (19)$$

And if $j = t-1$,

$$(y_t|\tau_t = t-1, s_t = s, k) \sim t_1 \left(2\alpha_1^{(k,s)}, X_1\mu_0^{(k,s)}, \frac{\alpha_2^{(k,s)}}{\alpha_1^{(k,s)}}\left(1 + X_1\Sigma_0^{(k,s)}X_1^T\right)\right) \tag{20}$$

The proof of Theorem 2 is given in Appendix C.

B. RUL Prediction

RUL prediction is to determine the time when the signal first hit the failure threshold Γ. For an operating unit, denote the remaining useful life as R_t at current time t. Then R_t can be defined as $R_t = \inf\{L: y_{t+L} \geq \Gamma|y_{1:t}\}$. The cumulative distribution function (CDF) of R_t conditional on available observations $y_{1:t}$ (also called conditional reliability function) can be expressed as

$$P(R_t > L|y_{1:t}) = \sum_k \sum_s \sum_j P(R_t > L|\tau_t = j, s_t = s, k, y_{1:t})P(\tau_t = j, s_t = s, k|y_{1:t}) \tag{21}$$

where $P(\tau_t, s_t, k|y_{1:t})$ is recursively calculated through Eq. (16) in the model updating stage, and $P(R_t > L|\tau_t, s_t, k, y_{1:t})$ can be reformulated as

$$P(R_t > L|\tau_t = j, s_t = s, k, y_{1:t}) = P(y_{t+1} < \Gamma, \dots, y_{t+L} < \Gamma|\tau_t = j, s_t = s, k, y_{1:t}) \tag{22}$$

The calculation of Eq. (22) depends on the segment index s_t. If $s_t = k + 1$, or the degradation process is at the final segment, it can be shown that the vector $y_{t+1:t+L}$ follows a multivariate t distribution of dimension L, as shown in Eq. (23) of Theorem 3.

Theorem 3 If $s_t = k + 1$,

$$(y_{t+1:t+L}|\tau_t = j, s_t = k+1, k, y_{j+1:t})$$
$$\sim t_L\left(2\alpha_1^{(k,k+1)} + t - j, X_{t+1-j,t+L-j}\mu_{j+1,t}, \frac{2\alpha_2^{(k,k+1)} + H_{j+1:t}^{(k,k+1)}}{2\alpha_1^{(k,k+1)} + t - j}(I + X_{t+1-j,t+L-j}\Sigma_{j+1,t}X_{t+1-j,t+L-j}^T)\right), \tag{23}$$

and

$$(y_{t+1:t+L}|\tau_{t+1} = j, s_{t+1} = k+1, k)$$
$$\sim t_L\left(2\alpha_1^{(k,k+1)}, X_{t+1-j,t+L-j}\mu_0^{(k,k+1)}, \frac{\alpha_2^{(k,k+1)}}{\alpha_1^{(k,k+1)}}(I + X_{t+1-j,t+L-j}\Sigma_0^{(k,k+1)}X_{t+1-j,t+L-j}^T)\right) \tag{24}$$

The proof of Theorem 3 is similar to Theorem 2 and thus is not provided here. Based on Theorem 3, if $s_t = k + 1$, i.e., the degradation is at the final stage, $P(R_t > L|\tau_t = j, s_t = k + 1, k, y_{1:t}) = \mathrm{MT}_{t+1:t+L}(\Gamma)$ where $\mathrm{MT}_{t+1:t+L}(\Gamma)$ is the CDF of

L-dimensional t distribution given in Eq. (23). If $s_t < k + 1$, however, the future change-points, especially the final change-point, need to be predicted for RUL prediction. The calculation for the general case is derived as follows. Denote the last or final change-point as c_k, then

$$P(R_t > L|\tau_t, s_t, k, y_{1:t}) = \sum_{c_k} P(R_t > L|c_k, \tau_t, s_t, k, y_{1:t})P(c_k|\tau_t, s_t, k, y_{1:t}) \tag{25}$$

where $P(c_k|\tau_t, s_t, k, y_{1:t})$ is the predictive PMF of the final change-point, which does not depend $y_{1:t}$ and thus can be recursively calculated in the off-line stage based on the defined Markov transition process in Eq. (18), and $P(R_t > L|c_k, \tau_t, s_t, k, y_{1:t})$ can be calculated based the model assumption that the degradation signal will not exceed the failure threshold before it reaches the final segment:

$$P(R_t > L|c_k, \tau_t, s_t, k, y_{1:t})$$
$$= \begin{cases} \mathrm{MT}_{t+1,t+L}(\Gamma), & \text{if } c_k = \tau_t \leq t - 1 \\ \mathrm{MT}_{c_k+1,t+L}(\Gamma), & \text{if } t - 1 < c_k < t + L \\ 1, & \text{if } c_k \geq t + L \end{cases} \tag{26}$$

In Eq. (26), $\mathrm{MT}_{t+1,t+L}(\Gamma)$ is the CDF of t distribution given in Eq. (23) while $\mathrm{MT}_{c_k+1,t+L}(\Gamma)$ is the CDF of $(y_{c_k+1:t+L}|\tau_{c_k+1} = c_k, s_{c_k+1} = k + 1, k)$ given in Eq. (24). As we can see, the closed form of the conditional reliability function or the RUL distribution can also be exactly obtained based on Eq. (21)-(26).

C. Computational Issue and Approximation

Although the model updating and RUL prediction can be exactly calculated through recursion, both the computational and memory cost of each time step increase with time t. From Eq. (16)-(18) we can see that the computational and memory cost of the filtering recursion at time t is approximately linear with time t, since we need to calculate and store $t\sum_k(k + 1)$ probabilities for the posterior PMF $P(\tau_t = j, s_t = s, k|y_{1:t})$. In the RUL prediction, for each L in Eq. (21), the computational cost also increases with t. For large datasets, these computational and storage issues may become very prohibitive in real time applications and thus need to be solved.

In practice, the posterior PMF $P(\tau_t = j, s_t = s, k|y_{1:t})$ is almost zero at most of the support points. Indeed, with more observations obtained in the current segment, the posterior PMF would concentrate around the starting point c_{s-1} of the current segment, and for $\tau_t \ll c_{s-1}$, the posterior PMF is close to zero and thus can be negligible. Similar phenomenon can also be observed for s_t and k. A natural way to control the computational cost and memory issue is to approximate the posterior densities at each time step with a small set of support points of fixed size that have high probabilities, and set the posterior PMF to be zero at the remaining support points. However, this strategy may result in inaccurate approximation if directly applied to the 3-dimensional support points. The

PMF can be temporally near zero for certain s_t and k, and then becomes dominant later as more observations are obtained. If the PMF is set zero at these support points, the PMF of these support points evaluated at the future time steps will also be zero, thus leading to an inaccurate approximation. To address this issue, we propose to select N most probable support points under each stratum (s_t, k) to approximate the posterior and set others to zero. Consequently, there are in total $N \sum_k (k + 1)$ non-zero support points. The details of the approximation algorithm are summarized in Table I.

TABLE I
SUMMARY OF THE APPROXIMATION UPDATING ALGORITHM

1. At time step $t = N + 1$
   - Calculate $P(\tau_t = j, s_t = s, k|y_{1:t})$ at all the $(N + 1)\sum_k(k + 1)$ support points.
   - Within each stratum (s, k), select N time steps from $\{1, \ldots, t\}$ with highest PMF $P(\tau_t = j, s, k|y_{1:t})$. Denote the selected time steps as $\boldsymbol{T}_N(t, s, k)$.
   - Normalize the probabilities of the selected $N\sum_k(k + 1)$ support points.

2. At time step $t > N + 1$
   - Calculate $P(\tau_t = j, s, k|y_{1:t})$ at $(N + 1)$ support points $\{\boldsymbol{T}_N(t - 1, s, k), t\}$ for each stratum (s, k).
   - Within each stratum (s, k), select N time steps from $\{\boldsymbol{T}_N(t - 1, s, k), t\}$ with highest PMF $P(\tau_t = j, s_t = s, k|y_{1:t})$. Update $\boldsymbol{T}_N(t, s, k)$.
   - Normalize the probabilities of the selected $N\sum_k(k + 1)$ support points.

Another computational issue is the calculation of the conditional reliability function through Eq. (21) and (25) at the online RUL prediction stage. It involves a large number of evaluations of the CDF of multivariate t distribution, e.g., $\text{MT}_{t+1,t+L}(\Gamma)$, whose computational cost increases enormously with the dimension L. To control the computational cost, we could alternatively use the Monte Carlo (MC) simulation approach. One MC approach is to directly generate the samples for the current state vector $\boldsymbol{x}_t$ through the posterior distribution $p(\boldsymbol{x}_t|y_{1:t})$ and simulate future state vectors $\boldsymbol{x}_{t+L}, L = 1, 2, \cdots$, through the prior Markov state transition process. Specifically, to simulate $\boldsymbol{x}_{t+1}$ conditioning on the previous state $\boldsymbol{x}_t$, we could first simulate the discrete components of $\boldsymbol{x}_{t+1}$ through Eq. (18), and then simulate the continuous component $\boldsymbol{\theta}_{t+1}$ from the prior distribution if t is a change-point or let $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t$ if t is not a change-point. Based on these simulated samples of current state and future states, the conditional reliability function can be easily calculated. However, due to the high dimensionality of the state vector, this MC approach requires a large number of samples to guarantee the approximation accuracy. To solve this issue, we propose to use partial MC simulation for only the calculation of $P(R_t > L|c_k, \tau_t, s_t, k, y_{1:t})$. Note that in Eq. (21) and (25), $P(c_k|\tau_t, s_t, k, y_{1:t})$ is independent of $y_{1:t}$ and can be calculated at the offline stage. $P(R_t > L|c_k, \tau_t, s_t, k, y_{1:t})$ is also independent of $y_{1:t}$ when $c_k \geq t$, and thus can also be calculated at the offline stage. Therefore we only need to estimate $P(R_t > L|c_k, \tau_t, s_t, k, y_{1:t})$ for $c_k = \tau_t \leq t - 1$ at the

online RUL prediction stage. To estimate it, we generate S samples $\{\boldsymbol{\theta}_{t1}, \ldots, \boldsymbol{\theta}_{tS}\}$ from the posterior distribution $P(\boldsymbol{\theta}_t|c_k, \tau_t, s_t, k, y_{1:t})$ given in Eq. (14), and then $P(R_t > L|c_k, \tau_t, s_t, k, y_{1:t})$ can be estimated by

$$
\begin{aligned}
&P(R_t > L|c_k, \tau_t, s_t, k, y_{1:t}) \\
&= \frac{1}{S} \sum_{i=1}^{S} \prod_{l=1}^{L} \Phi(\Gamma|\boldsymbol{X}_{t+l-c_k}\boldsymbol{\beta}_{ti}, \sigma_{ti}^2)
\end{aligned}
\tag{27}
$$

where $\Phi(\cdot |\boldsymbol{X}_{t+l-c_k}\boldsymbol{\beta}_{ti}, \sigma_{ti}^2)$ is the CDF of Gaussian distribution with mean $\boldsymbol{X}_{t+l-c_k}\boldsymbol{\beta}_{ti}$ and variance σ_{ti}^2. This strategy can significantly improve the computational efficiency yet without influencing the calculation accuracy. Another advantage of this strategy is that we can easily adopt truncated distributions for the last segment (e.g., only select samples with positive degradation rate) to avoid the occurrence of a temporary decreasing trend (unit would never fail and RUL prediction is infeasible) in the updated signal evolution path for RUL prediction, which is common in real degradation signals due to measurement noises [34]. Note that in partial MC method, only $P(R_t > L|c_k, \tau_t, s_t, k, y_{1:t})$ with $c_k = \tau_t \leq t - 1$ (i.e., degradation is at the final phase) is calculated by MC method. The approximation accuracy decreases with L for a fixed sample size. However, in many practical applications, the final phase is often very steep and short in duration, so that $P(R_t > L|c_k, \tau_t, s_t, k, y_{1:t})$ converges rapidly to 0 as L increases. Therefore, using a small sample size can often achieve a very accurate approximation.

V. CASE STUDIES

In this section, we first use simulated signals to illustrate the effectiveness of the proposed method, and then apply the method to vibrational signals of rotational bearings for performance evaluation.

A. Simulation Study

TABLE II
HYPERPARAMETERS FOR THE BAYESIAN MULTIPLE CHANGE-POINT MODEL

| Variables	$k = 1$	$k = 2$
$\delta^{(k,s)}$	$\delta_0^{(1,1)} = 500, \sigma_0^{2(1,1)} = 30^2$ $\delta_0^{(1,2)} = 500, \sigma_0^{2(1,2)} = 30^2$	$\delta_0^{(2,1)} = 500, \sigma_0^{2(2,1)} = 50^2$ $\delta_0^{(2,2)} = 400, \sigma_0^{2(2,2)} = 30^2$ $\delta_0^{(2,3)} = 500, \sigma_0^{2(2,3)} = 30^2$
$\boldsymbol{\beta}^{(k,s)}$	$\boldsymbol{\mu}_0^{(1,1)} = [-10; 0.005]$ $\boldsymbol{\Sigma}_0^{(1,1)} = \begin{bmatrix} 0.2 & 0.0015 \\ 0.0015 & 0.0008 \end{bmatrix}$ $\boldsymbol{\mu}_0^{(1,2)} = [-40; 0.05]$ $\boldsymbol{\Sigma}_0^{(1,2)} = \begin{bmatrix} 0.5 & -0.0008 \\ -0.0008 & 0.006 \end{bmatrix}$	$\boldsymbol{\mu}_0^{(2,1)} = [-10; 0.003]$ $\boldsymbol{\Sigma}_0^{(2,1)} = \begin{bmatrix} 0.15 & 0.0014 \\ 0.0014 & 0.0009 \end{bmatrix}$ $\boldsymbol{\mu}_0^{(2,2)} = [-20; 0.02]$ $\boldsymbol{\Sigma}_0^{(2,2)} = \begin{bmatrix} 0.024 & -0.0009 \\ -0.0009 & 0.000055 \end{bmatrix}$ $\boldsymbol{\mu}_0^{(2,3)} = [-30; 0.09]$ $\boldsymbol{\Sigma}_0^{(2,3)} = \begin{bmatrix} 0.75 & -0.00008 \\ -0.00008 & 0.00045 \end{bmatrix}$
$\sigma^{2(k,s)}$	$\alpha_1^{(1,1)} = 2, \alpha_2^{(1,1)} = 4$ $\alpha_1^{(1,2)} = 1, \alpha_2^{(1,2)} = 4$	$\alpha_1^{(2,1)} = 3.6, \alpha_2^{(2,1)} = 3$ $\alpha_1^{(2,2)} = 3, \alpha_2^{(2,2)} = 5$ $\alpha_1^{(2,3)} = 3.6, \alpha_2^{(2,3)} = 5$
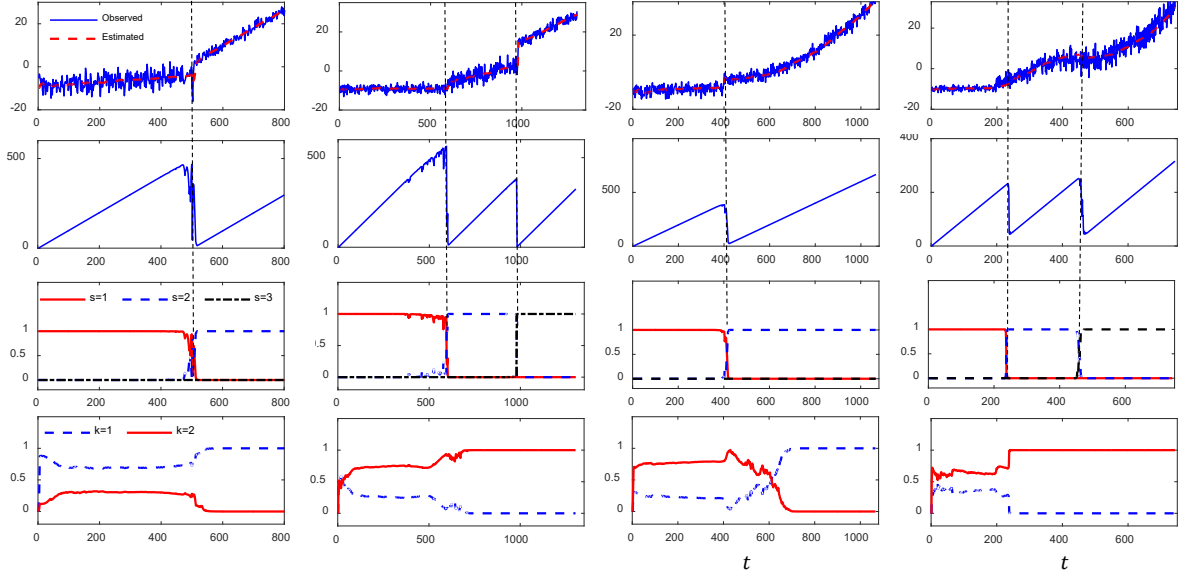
Fig. 3. Illustration of the online monitoring and model updating for two models. (a) and (b): all segments are simple linear models; (c) and (d): all the last segments are quadratic; first row: raw and estimated or filtered CM signals; second row: the expected duration of the current segment; third row: the posterior PMF of the index of current segment; and bottom row: the posterior PMF of the total number of segments (or signal type). The vertical dashed lines are true change-points.

To demonstrate the effectiveness of the proposed algorithm, a simulation study is conducted. In the simulation model, we assume that there are two types of degradation signals in terms of the number of change-points, namely, one-change-point (one-CP) and two-change-points (two-CP) signals. The corresponding probabilities are given by

$$k = \begin{cases} 1, \text{with } p_1 = 0.2 \\ 2, \text{with } p_2 = 0.8 \end{cases}$$
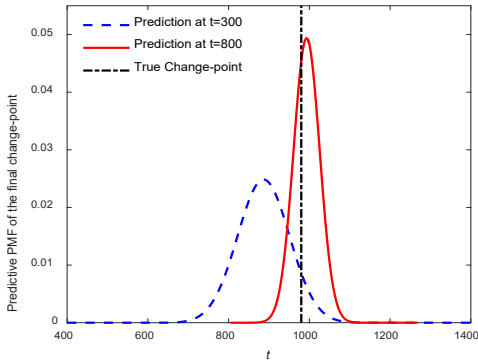

Fig. 4. Predictive PMF of the final change-point with prediction time $t = 300$ and $t = 800$.

All the segments are modelled as simple linear regression or line segment for both one-CP and two-CP signals. The other hyperparameters of the Bayesian multiple change-point model are given in Table II. In total $I = 300$ signals are simulated as the historical dataset. The failure threshold is set as $\Gamma = 30$. Based on simulation settings above, 64 signals are generated for $k = 1$ and 236 signals are generated for $k = 2$. The hyperparameters are then estimated by the empirical two-stage estimation approach introduced in Section 3.2. The proposed

method can detect the change points (the number and positions) accurately for each signal. Due to page limitation, the estimated hyperparameters are not listed here. For illustration, we also consider another model where all the settings are the same with the model described above except that all the last segments are quadratic. The hyperparameters of this model are not provided here due to space limitation.
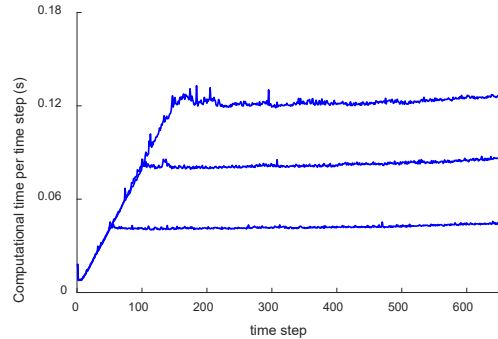

Fig. 5. Computational time per time step for three different support sizes N in the model updating process.

Fig. 3 illustrates the online condition monitoring and model updating of individual unit for these two degradation models. Unless otherwise specified, the support size $N = 5$ for each stratum (s_t, k) for posterior approximation in this paper. Clearly, the sequentially calculated posterior distribution of the discrete components of the state vector could effectively detect the occurrence of change-points, track the index of the current segment, and infer how many segments the CM signal would have. The accurate estimation of the CM signals also indicates an effective updating of the posterior distribution of the continuous components or model parameters of each segment. Fig. 4 shows an example of predicting the position of the final
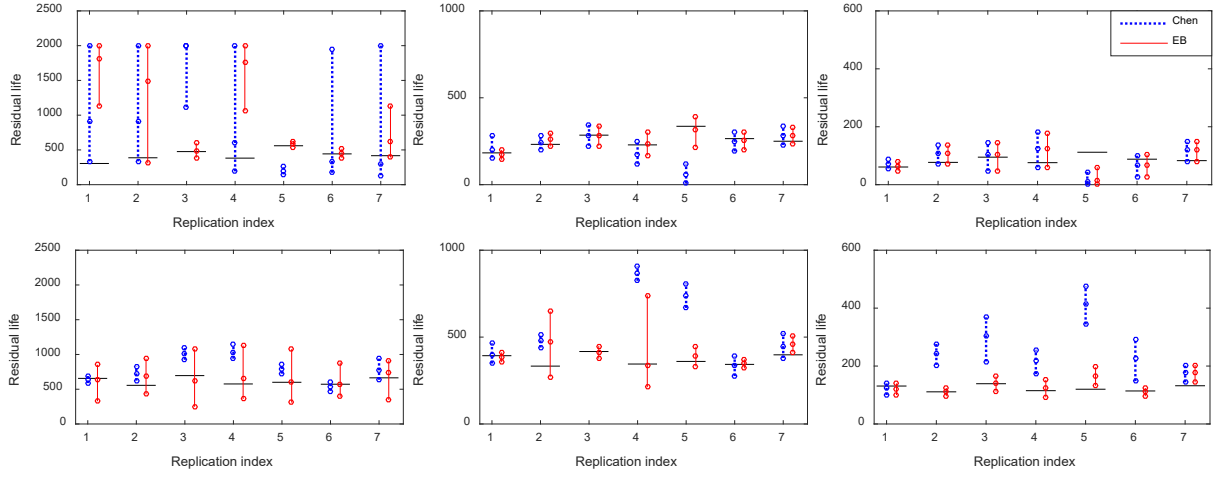
Fig. 6. Prediction intervals for 7 one-CP signals (a, b and c) and 7 two-CP signals (d, e and f) at three different prediction times. (a) and (d): 50% of failure time; (b) and (e): 70% of failure time; (c) and (f): 90% of failure time. The ∘ represents the 5%, 50%, 95% quantiles of the predicted RUL distributions, and − denotes the actual RUL.
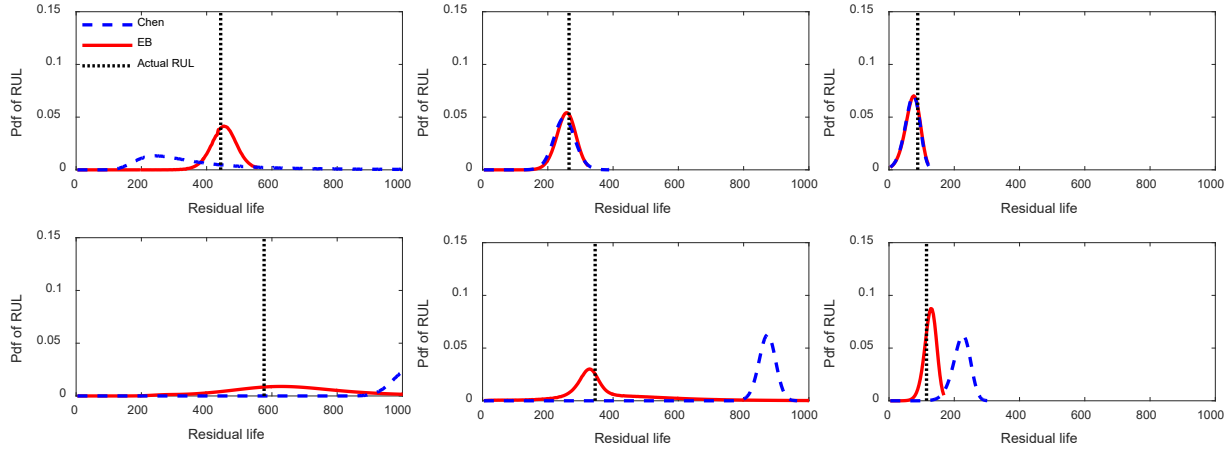

Fig. 7. Comparison of the detailed pdf of the RUL. (a), (b) and (c) correspond to the 6th signal (one-CP) of Fig. 6(a), (b) and (c), respectively; (d), (e) and (f) correspond to the 4th signal (two-CP) of Fig. 6(d), (e) and (f) respectively.

change-point at two different times using the degradation signal shown in Fig. 3(b). At $t = 300$, the degradation signal is within its first segment. Hence, it is high likely that there are two future change-points to be predict, which increases the prediction uncertainty. At $t = 800$, the degradation signal has already transited to the second segment, and the final change-point can be accurately predicted. Clearly, as more data are observed, the prediction of the final change-point becomes more accurate.

Fig. 5 shows the computational cost of each time step in the posterior model updating process using three different support sizes $N = 50, 100$ and 150. As we can clearly see, when $t < N$, the computational cost linearly increases with t, which is consistent with what we discussed in Section 4.3. Once the approximation strategy with a fixed support size N is applied, the computational cost of each step is fully controlled for $t \geq N$ with an almost constant computational time.

To evaluate the performance of RUL prediction, another 100 CM signals are simulated as testing dataset. The proposed method (denote it as EB) is compared with Chen's method [13], which models CM signals with two line segments and thus is an

ideal method for comparison. For Chen's method, all 300 training CM signals (236 signals with three line segments and 64 with two line segments) are used to estimate the hyperparameters of the two line segments. For the EB method, the support size $N = 5$.

Fig. 6 shows the prediction intervals of EB and Chen's method at three prediction times for 7 one-CP signals and 7 two-CP signals randomly selected from the testing dataset. Fig. 7 shows the detailed pdf of the predicted RUL for the 6th signal of Fig. 6(a-c) and the 4th signal of Fig. 6 (d-f). Unsurprisingly, the prediction for both methods becomes more and more accurate as more observations are available. Comparing these two methods, the proposed EB method outperforms Chen's method for almost all the 14 signals. For one-CP signals, the EB method is slightly better at 70% and 90% of failure time, while at the prediction time 50%, the advantage of EB method is much more significant. The reason is that at the early degradation stage, the priors play a decisive role on the prediction accuracy. In Chen's method, the priors of two-line-segment model are estimated using all one-CP and two-CP signals, which results in inaccurate priors. At 70% and

90% of failure time, all degradation signals evolve into the second segment, and the posterior distribution of each model is dominated by the observations. Therefore, Chen's method can also accurately predict the RUL with inaccurate priors. For two-CP signals, the EB method is much better than Chen's method, which fulfills the purpose of the proposed method for degradation signals with two or even more segments.

To better quantify the prediction performance, the $\alpha - \lambda$ performance metric [35] is calculated, where α specifies the error bound on the estimated RUL, i.e., $[(1 - \alpha)]R_{i,true} \leq \hat{R}_i \leq [1 + \alpha]R_{i,true}$, and λ specifies the relative distance, in time, of a given prediction point from the actual failure time, i.e., $\lambda = 0$ and $\lambda = 1$ correspond to the starting prediction time and the actual failure time, respectively. Fig. 8 shows the $\alpha - \lambda$ performance metric for the 6th one-CP signal and 4th two-CP signal using the proposed EB method and Chen's method. The error bound α is set as 20%. It can be observed that, almost all the estimated RULs lie within the error bound for one-change-point case using both methods. However, for two-change-points case, Chen's method is much worse than the proposed EB method.
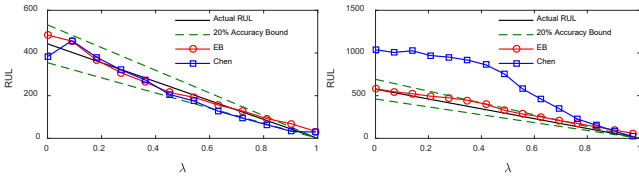

Fig. 8. $\alpha - \lambda$ performance metric for (a) the 6th signal for one-CP case, and (b) the 4th signal for two-CP case.

TABLE III
COMPARISON OF THE RMSD AT SIX PREDICTION TIMES

Method	RMSD					
	40%	50%	60%	70%	80%	90%
Chen	319.5	271.6	218.3	240.6	211.0	102.0
EB	313.8	188.8	153.8	114.4	35.1	23.0
EB-partial MC	318.0	203.7	149.6	103.8	37.7	31.3

To evaluate the overall performance of RUL prediction, we use the root-mean-square-deviation (RMSD) for these 100 testing signals defined as

$$\text{RMSD} = \sqrt{\frac{1}{100}\sum_{i=1}^{100} E\left(\hat{R}_i - R_{i,\text{true}}\right)^2}$$

where $\hat{R}_i$ and $R_{i,\text{true}}$ are the predicted true RUL of unit i, respectively. Table III shows the RMSD of the proposed EB method and Chen's method at six prediction times. We also add the EB method with partial Monte Carlo approximation (denoted as EB-partial MC) to see how partial MC influence the prediction accuracy. For EB-partial MC, a sampling size of 1000 is used. As we can see, the proposed EB method is much more accurate than Chen's method. At the early stage, e.g., 40% of the failure time, the advantage of the EB method is not significant, due to large uncertainty of model parameters and future change-points. As the prediction time approaches to the

true failure time, the RMSD of the proposed method decreases significantly, and the performance is much better than Chen's method. This is highly desirable since it becomes more and more important to get an accurate prediction when the RUL approaches zero. Comparing EB with EB-partial MC we can see that the prediction accuracy is not influenced much by partial MC sampling strategy.

Table IV shows the computational costs of the EB method and EB-partial MC using MATLAB running on an i5-4690 CPU 3.50 GHz Intel processor at the prediction stage. In the RUL prediction, the computational times are calculated under different prediction steps. For example, if the prediction step is L, the conditional survival function $P(R_t > l|y_{1:t})$ is evaluated for $l = 1,2,\cdots,L$, with in total L calculations. As we can see, the cost of EB method exponentially increases with the prediction step, due to the CDF computation of multivariate t distributions with increasing dimensions. For the EB-partial MC method, the computational cost of the prediction linearly increases with the prediction step. Therefore, using the partial MC strategy for the EB method could significantly reduce the computational cost, yet without influencing the prediction accuracy much.

TABLE IV
THE COMPUTATIONAL COST OF RUL PREDICTION WITH AND WITHOUT PARTIAL MONTE CARLO SIMULATION (UNIT: SECONDS)

Method	Prediction Steps					
	10	20	30	40	50	60
EB	0.8	4.0	8.8	14.6	21.8	30.4
EB-partial MC	4.7	5.2	5.7	6.3	6.8	7.4

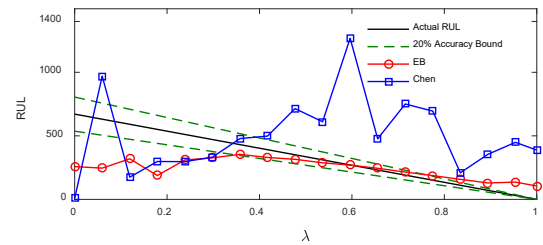### B. Application to Rotational Bearings

Fig. 9. $\alpha - \lambda$ performance metric for the 24th bearing signal

In this section, we apply the proposed EB method to the real degradation signals of rolling thrust bearings [8, 12, 13]. To generate these signals, a set of identical thrust bearings was run at a constant rotational speed (2000 r/min) and a load of 200 lbs in an oil bath to provide continuous lubrication. Then the vibrations frequencies were acquired from an accelerometer, which was attached to the setup and connected to a vibration meter that measured the rms vibration level [36]. The amplitude of these frequencies increases as the bearing degrades. The degradation signal used in this paper consists of the average amplitude of the defective frequency and its first six harmonics frequencies. The degradation amplitudes are log-transformed. As the bearing degrades, the vibration becomes more and more severe and thus the degradation signal tends to increase. When the vibration magnitude reaches a threshold, the bearing is

TABLE V
ESTIMATED HYPERPARAMETERS OF THE PRIOR DISTRIBUTIONS

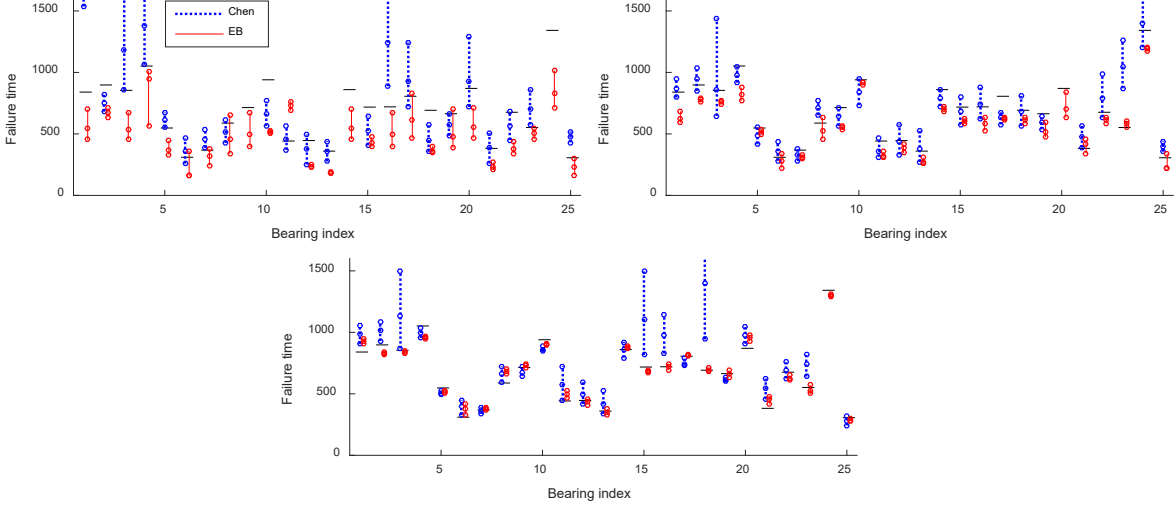|	$s = 1$	$s = 2$	$s = 3$
$\delta^{(s)}$	$\delta_0^{(1)} = 246,\ \sigma_0^{2(1)} = 163^2$	$\delta_0^{(2)} = 199,\ \sigma_0^{2(2)} = 123^2$	$\delta_0^{(3)} = 232,\ \sigma_0^{2(3)} = 123^2$
$\boldsymbol{\beta}^{(s)}$	$\boldsymbol{\mu}_0^{(1)} = [-7.28, 5.6 \times 10^{-6}]$ $\boldsymbol{\Sigma}_0^{(1)} = \begin{bmatrix} 20.77 & -0.072 \\ -0.072 & 0.0020 \end{bmatrix}$	$\boldsymbol{\mu}_0^{(2)} = [-6.97, 0.0015]$ $\boldsymbol{\Sigma}_0^{(2)} = \begin{bmatrix} 54.79 & -0.075 \\ -0.075 & 0.076 \end{bmatrix}$	$\boldsymbol{\mu}_0^{(3)} = [-4.76, 0.004]$ $\boldsymbol{\Sigma}_0^{(3)} = \begin{bmatrix} 3.74 & -0.0004 \\ -0.0004 & 0.0002 \end{bmatrix}$
$\sigma^{2(s)}$	$\alpha_1^{(1)} = 5.17,\ \alpha_2^{(1)} = 0.03$	$\alpha_1^{(2)} = 1.06,\ \alpha_2^{(2)} = 0.008$	$\alpha_1^{(3)} = 8.7,\ \alpha_2^{(3)} = 0.38$

Fig. 10. Prediction intervals for 25 bearing signals at three different prediction times. (a): 50%; (b): 70%; and (c): 90% of the failure time. The ∘ represents the 5%, 50%, 95% quantiles of the predicted RUL distributions, and – denotes the actual failure time.
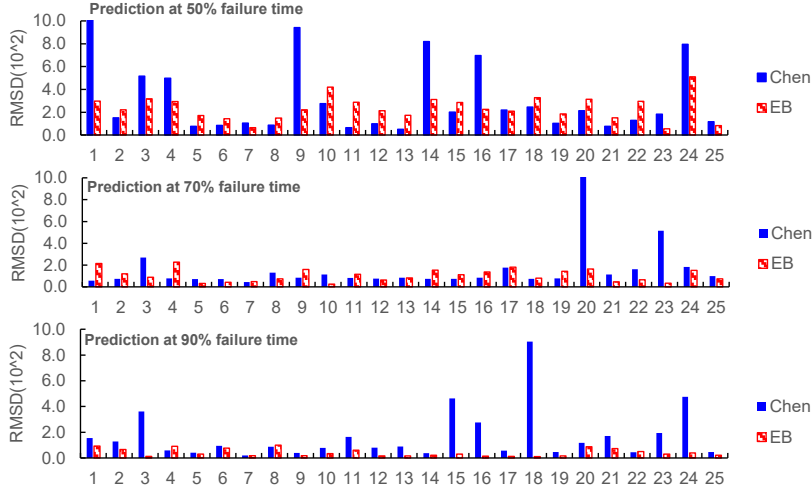

Fig. 11. RMSD of 25 bearing signals

considered to have failed. Based on the published industrial standards, the failure threshold is set $\Gamma = \log 0.03$ [8]. In total there are 25 historical signals. The sampling interval for all signals is 2 minutes.

In the offline modeling, we set the maximum number of change-points to be 2 to control the model complexity. We assume that all segments are line segments. Based on the BIC model selection, two-CP model is the best for all signals. The estimated hyperparameters are summarized in Table V. The estimated means of three slopes show that degradation rate is almost zero at the first stage, indicating a stable operation, and then increases successively at the following two stages.

Fig. 9 shows the α-λ performance metric for the 24th bearing signal. It can be observed that, although some estimated values are outside of the accuracy bound at early stage, the proposed method makes quite accurate prediction at later prediction stage. Apparently, the proposed method has a better performance. Fig. 10 shows the prediction intervals at 50%, 70% and 90% of failure time for the 25 degradation signals. The prediction results of Chen's method is also provided for comparison. Some intervals by Chen's method are not shown

since they are out of the y-axis range, e.g., the 9th and 14th signal on Fig. 10 (a), the 20th signal on Fig. 10 (b) and the 24th signal on Fig. 10 (c). Fig. 11 shows the RMSD of the 25 bearing signals for both EB and Chen's method. Clearly, as the prediction time is closer to the failure time, the intervals become narrower. The more observed data, the more accurate the prediction. The prediction of the EB method is very stable and accurate across all units, while Chen's method is not stable and performs badly for some units. Table VI shows the comparison of the proposed EB method with Chen's method, the GLLR method [8], the GBPNN method [37] and EB method using three change-points (EB-CP3) in terms of the overall RMSD at the three time steps. In the GLLR method, the remaining data was fitted using Bayesian simple linear regression. In the GBPNN method, a back-propagation neural network-based model was developed for RUL prediction. It is noted that, in the GLLR and GBPNN methods, the first segment with normal working condition is manually truncated. Therefore, we just show the prediction at 70% and 90% prediction time for GLLR and GBPNN methods. Clearly, the proposed method outperforms all of other methods at all three prediction times. It should be mentioned that EB-CP3 is even worse. The reason is that most of bearings just have 1 or 2 obvious change-points, adding excessive change points will also introduce unnecessary uncertainty in RUL prediction, i.e., uncertainty of future change-point locations. Therefore, the number of change-points is critical for the prognostic model to generate an accurate prediction.

TABLE VI

RMSD OF THE PROPOSED METHOD IN COMPARISON WITH OTHER METHODS

Method	RMSD		
	50%	70%	90%
GLLR	-	234.2	227.8
GBPNN	-	193.2	174.1
Chen	318.4	156.9	169.4
EB	236.8	106.0	41.8
EB-CP3	493.9	297.1	179.2

VI. CONCLUSION AND DISCUSSION

In this paper we proposed a Bayesian multiple change-point modeling framework for degradation signals based condition monitoring and remaining useful life prediction. To capture the unit-to-unit heterogeneity and also to facilitate integration of historical data with in-situ observations of in-service unit for online prognostics, all model parameters are assumed to be random, including the number of change-points and their positions, and the model parameters of each linear segments. A novel stochastic process was proposed to model the joint prior of change-points and positions. A two-stage process was proposed to estimate all hyperparameters of priors. To facilitate online Bayesian model updating, a recursive updating algorithm was developed by which the posterior distribution of all state parameters can be exactly calculated. A closed-form of the RUL prediction is also derived. To control the computational cost in both model updating and RUL prediction process, a fixed-support-size strategy and a partial Monte Carlo strategy were proposed respectively, which significantly reduced the computational cost without influencing the prediction accuracy. The advantages of the proposed method have been demonstrated through thorough simulation studies and real case studies.

There are still open questions worthy of investigation. First of all, the current multiple change-point model assumes that all segments are independent. However, the degradation signals are often continuous in practice, indicating that all segments are connected and dependent. Incorporating such dependence could make the prior more informative and thus improve the prognostic accuracy. Secondly, adding more change-points may improve the model fitting and improve the prediction accuracy at the late degradation stage. However, it may reduce the prediction accuracy at the early degradation stage due to extra uncertainty by the added change-points. The strategy of using different models at different prediction stage may be beneficial. Lastly, the segments of the multiple change-point model are modeled by parametric regression. It would be of interest to integrate the multiple change-point model to stochastic processes, e.g., Wiener process, for degradation modeling.

APPENDIX A DERIVATION OF EQUATION (12)

For notational convenience, we ignore the superscript k and s here. Suppose $\{\boldsymbol{\beta}_i, \sigma_i^2\}_{i=1}^{i=n}$ are the observed samples from distribution

$$\pi(\boldsymbol{\beta}, \sigma^2) = IG(\sigma^2|\alpha_1, \alpha_2)N(\boldsymbol{\mu_0}, \sigma^2\boldsymbol{\Sigma_0})$$

The likelihood function can be written as

$$L(\alpha_1, \alpha_2, \boldsymbol{\mu_0}, \boldsymbol{\Sigma_0}|\{\boldsymbol{\beta}_i, \sigma_i^2\}_{i=1}^{i=n})$$
$$\propto \prod_{i=1}^{n}\left[\frac{\alpha_2^{\alpha_1}}{\Gamma(\alpha_1)}(\sigma_i^2)^{-\alpha_1-1}\exp\left(-\frac{\alpha_2}{\sigma_i^2}\right)\right]|\sigma_i^2\boldsymbol{\Sigma_0}|^{-\frac{1}{2}}\exp\left[-\frac{(\boldsymbol{\beta}_i-\boldsymbol{\mu_0})'\boldsymbol{\Sigma_0}^{-1}(\boldsymbol{\beta}_i-\boldsymbol{\mu_0})}{2\sigma_i^2}\right]$$

The log-likelihood function l is

$$l\left(\alpha_1, \alpha_2, \boldsymbol{\mu_0}, \boldsymbol{\Sigma_0}\Big|\{\boldsymbol{\beta}_i, \sigma_i^2\}_{i=1}^{i=n}\right)$$
$$= \sum_{i=1}^{n}\log\left[\frac{\alpha_2^{\alpha_1}}{\Gamma(\alpha_1)}(\sigma_i^2)^{-\alpha_1-1}\exp\left(-\frac{\alpha_2}{\sigma_i^2}\right)\right]$$
$$- \frac{n}{2}\log|\boldsymbol{\Sigma_0}| - \sum_{i=1}^{n}\frac{(\boldsymbol{\beta}_i-\boldsymbol{\mu_0})'\boldsymbol{\Sigma_0}^{-1}(\boldsymbol{\beta}_i-\boldsymbol{\mu_0})}{2\sigma_i^2}$$
$$+ C$$

Therefore the MLE of α_1, α_2 can be obtained by maximizing the log-likelihood function $l(\alpha_1, \alpha_2|\{\sigma_i^2\}_{i=1}^{i=n})$. For $\boldsymbol{\mu_0}, \boldsymbol{\Sigma_0}$, the log-likelihood function is

$$l(\boldsymbol{\mu_0}, \boldsymbol{\Sigma_0}|\{\boldsymbol{\beta}_i, \sigma_i^2\}_{i=1}^{i=n}) = -\frac{n}{2}\log|\boldsymbol{\Sigma_0}| - \sum_{i=1}^{n}\frac{(\boldsymbol{\beta}_i-\boldsymbol{\mu_0})'\boldsymbol{\Sigma_0}^{-1}(\boldsymbol{\beta}_i-\boldsymbol{\mu_0})}{2\sigma_i^2} + C$$
$$\frac{\partial l}{\partial \boldsymbol{\mu_0}} = -\sum_{i=1}^{n}\frac{\boldsymbol{\Sigma_0}^{-1}(\boldsymbol{\mu_0}-\boldsymbol{\beta}_i)}{\sigma_i^2} = 0$$

Therefore

$$\hat{\boldsymbol{\mu}}_0 = \sum_{i=1}^{n}\frac{\boldsymbol{\beta}_i}{\sigma_i^2}\Bigg/\sum_{i=1}^{n}\frac{1}{\sigma_i^2}$$

Plug in $\hat{\boldsymbol{\mu}}_0$ we can get

$$l(\hat{\boldsymbol{\mu}}_0, \boldsymbol{\Sigma_0}|\{\boldsymbol{\beta}_i, \sigma_i^2\}_{i=1}^{i=n})$$
$$= -\frac{n}{2}\log|\boldsymbol{\Sigma_0}| - \sum_{i=1}^{n}\frac{(\boldsymbol{\beta}_i-\hat{\boldsymbol{\mu}}_0)'\boldsymbol{\Sigma_0}^{-1}(\boldsymbol{\beta}_i-\hat{\boldsymbol{\mu}}_0)}{2\sigma_i^2} + C$$
$$= -\frac{n}{2}\log|\boldsymbol{\Sigma_0}| - \frac{1}{2}\text{tr}\left(\sum_{i=1}^{n}\frac{\boldsymbol{\Sigma_0}^{-1}(\boldsymbol{\beta}_i-\hat{\boldsymbol{\mu}}_0)(\boldsymbol{\beta}_i-\hat{\boldsymbol{\mu}}_0)'}{\sigma_i^2}\right)$$

$$= -\frac{n}{2}\log|\boldsymbol{\Sigma_0}| - \frac{n}{2}\text{tr}(\boldsymbol{\Sigma_0}^{-1}S)$$

where

$$S = \frac{1}{n}\sum_{i=1}^{n}\frac{(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\mu}}_0)(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\mu}}_0)'}{\sigma_i^2}$$

It can be shown that $\widehat{\boldsymbol{\Sigma}}_0 = S$.

Let $S = EE^T, \Psi = E^T\boldsymbol{\Sigma_0}^{-1}E$, then $|\boldsymbol{\Sigma_0}| = \frac{|S|}{|\Psi|}$, $\text{tr}(\boldsymbol{\Sigma_0}^{-1}S) = \text{tr}(\boldsymbol{\Sigma_0}^{-1}EE^T) = \text{tr}(E^T\boldsymbol{\Sigma_0}^{-1}E) = \text{tr}(\Psi)$.

So $l(\boldsymbol{\Sigma_0}, \widehat{\boldsymbol{\mu}}_0|\{\boldsymbol{\beta}_i, \sigma_i^2\}_{i=1}^{i=n}) = -\frac{n}{2}(\log|S| - \log|\Psi| + \text{tr}(\Psi))$

Let $\Psi = LL^T$ where L is a lower triangular matrix

$$l(\boldsymbol{\Sigma_0}, \widehat{\boldsymbol{\mu}}_0|\{\boldsymbol{\beta}_i, \sigma_i^2\}_{i=1}^{i=n})$$
$$= -\frac{n}{2}\left(\log|S| - \sum_{i=1}^{2}\log l_{ii}^2 + \sum_{i=1}^{2}l_{ii}^2 + \sum_{i>j}l_{ij}^2\right)$$

It is easy to show that when $l_{ij} = 0$ for $i \neq j$ and $l_{ii}^2 = 1$, $l(\boldsymbol{\Sigma_0}, \widehat{\boldsymbol{\mu}}_0|\{\boldsymbol{\beta}_i, \sigma_i^2\}_{i=1}^{i=n})$ reaches the maximum

Therefore $\widehat{\Psi} = I, \widehat{\Sigma}_0 = S$

APPENDIX B PROOF OF THEOREM 1

For notational convenience, we ignore the superscripts k and s here. We also ignore the subscript t for $\boldsymbol{\beta}_t$ and σ_t^2. Suppose $\pi(\boldsymbol{\beta}, \sigma^2) = IG(\sigma^2|\alpha_1, \alpha_2)N(\boldsymbol{\mu_0}, \sigma^2\boldsymbol{\Sigma_0})$ and $\boldsymbol{\beta}$ is of dimension q.

$$p(\boldsymbol{\beta}, \sigma^2|y_{j+1:t}) \propto p(\boldsymbol{\beta}, \sigma^2)p(y_{j+1:t}|\boldsymbol{\beta}, \sigma^2)$$
$$\propto \left[\frac{1}{2\pi|\sigma^2\boldsymbol{\Sigma_0}|^{\frac{1}{2}}}e^{-\frac{(\boldsymbol{\beta}-\boldsymbol{\mu_0})^T\boldsymbol{\Sigma_0}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu_0})}{2\sigma^2}}\right]\left[\frac{\alpha_2^{\alpha_1}}{\Gamma(\alpha_1)}(\sigma^2)^{-\alpha_1-1}e^{-\frac{\alpha_2}{\sigma^2}}\right]$$
$$\left[(2\pi)^{-\frac{t-j}{2}}(\sigma^2)^{-\frac{t-j}{2}}e^{\frac{-\|y_{j+1:t}-X_{1,t-j}\boldsymbol{\beta}\|^2}{2\sigma^2}}\right]$$
$$\propto (\sigma^2)^{-\alpha_1-1-\frac{t-j}{2}}\frac{1}{2\pi(\sigma^2)^{\frac{q}{2}}}$$
$$\exp\left[-\frac{(\boldsymbol{\beta}-\boldsymbol{\mu_0})^T\boldsymbol{\Sigma_0}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu_0}) + 2\alpha_2 + \|y_{j+1:t} - X_{1:t-j}\boldsymbol{\beta}\|^2}{2\sigma^2}\right]$$
$$\propto (\sigma^2)^{-\alpha_1-\frac{(t-j)}{2}-1}$$
$$\exp\left[-\frac{y_{j+1:t}^Ty_{j+1:t} + 2\alpha_2 + \boldsymbol{\mu_0}^T\boldsymbol{\Sigma_0}^{-1}\boldsymbol{\mu_0} - \boldsymbol{\mu}_{j+1:t}^T(X_{1:t-j}^TX_{1:t-j}+\boldsymbol{\Sigma_0}^{-1})\boldsymbol{\mu}_{j+1:t}}{2\sigma^2}\right]$$
$$\times\frac{1}{(\sigma^2)^{q/2}}\exp\left[-\frac{(\boldsymbol{\beta}-\boldsymbol{\mu}_{j+1:t})^T(X_{1:t-j}^TX_{1:t-j}+\boldsymbol{\Sigma_0}^{-1})(\boldsymbol{\beta}-\boldsymbol{\mu}_{j+1:t})}{2\sigma^2}\right]$$
$$\propto IG\left(\alpha_1 + \frac{t-j}{2}, \alpha_2 + \frac{H_{j+1,t}}{2}\right)\cdot N(\boldsymbol{\mu}_{j+1,t}, \sigma^2\boldsymbol{\Sigma}_{j+1,t})$$

where

$$\boldsymbol{\Sigma}_{j+1,t} = (X_{1,t-j}^TX_{1,t-j} + \boldsymbol{\Sigma_0}^{-1})^{-1}$$
$$\boldsymbol{N}_{j+1,t} = (\boldsymbol{\Sigma_0}^{-1}\boldsymbol{\mu_0}^{(k,s)} + X_{1,t-j}^Ty_{j+1:t})$$
$$\boldsymbol{\mu}_{j+1,t} = \boldsymbol{\Sigma}_{j+1,t}\boldsymbol{N}_{j+1,t}$$
$$H_{j+1,t} = y_{j+1:t}^Ty_{j+1:t} + \boldsymbol{\mu_0}^T\boldsymbol{\Sigma_0}^{-1}\boldsymbol{\mu_0} - \boldsymbol{N}_{j+1,t}^T\boldsymbol{\Sigma}_{j+1,t}\boldsymbol{N}_{j+1,t}$$

APPENDIX C PROOF OF THEOREM 2

If $\tau_t = j < t - 1$, based on Theorem 1 we can get

$$(\boldsymbol{\beta}_{t-1}|\sigma_{t-1}^2, \tau_{t-1} = j, s_{t-1} = s, k, y_{j+1:t-1}) \sim N(\boldsymbol{\mu}_{j+1,t-1}, \sigma_{t-1}^2\boldsymbol{\Sigma}_{j+1,t-1})$$
$$(X_{t-j}\boldsymbol{\beta}_{t-1}|\sigma_{t-1}^2, \tau_{t-1} = j, s_{t-1} = s, k, y_{j+1:t-1}) \sim$$

$$N(X_{t-j}\boldsymbol{\mu}_{j+1,t-1}, \sigma_{t-1}^2X_{t-j}\boldsymbol{\Sigma}_{j+1,t-1}X_{t-j}^T)$$

$$(\sigma_{t-1}^2|\tau_{t-1} = j, s_{t-1} = s, k, y_{j+1:t-1}) \sim IG\left(\alpha_1^{(k,s)} + \frac{t-1-j}{2}, \alpha_2^{(k,s)}\right.$$
$$\left. + \frac{H_{j+1,t-1}}{2}\right)$$

Since $y_t = X_{t-j}\boldsymbol{\beta}_{t-1} + \sigma_{t-1}\varepsilon_t$ for $\tau_t = \tau_{t-1} = j < t - 1$, then

$$(y_t|\sigma_{t-1}^2, \tau_{t-1} = j, s_{t-1} = k, y_{j+1:t-1}) \sim N(X_{t-j}\boldsymbol{\mu}_{j+1,t-1}, \sigma_{t-1}^2(1 + X_{t-j}\boldsymbol{\Sigma}_{j+1,t-1}X_{t-j}^T)),$$

Let $\mu_* = X_{t-j}\boldsymbol{\mu}_{j+1,t-1}, \sigma_*^2 = 1 + X_{t-j}\boldsymbol{\Sigma}_{j+1,t-1}X_{t-j}^T$

$$p(y_t|\tau_t = j, s_t = s, k, y_{1:t-1})$$
$$= \int p(y_t|\sigma_{t-1}^2, \tau_{t-1} = j, s_{t-1} = s, k, y_{1:t-1})$$
$$p(\sigma_{t-1}^2|\tau_{t-1} = j, s_{t-1} = s, k, y_{1:t-1})d\sigma_{t-1}^2$$
$$\propto \int (\sigma_{t-1}^2\sigma_*^2)^{-\frac{1}{2}}\exp\left[-\frac{(y_t-\mu_*)^2}{2\sigma_{t-1}^2\sigma_*^2}\right](\sigma_{t-1}^2)^{-\alpha_1^{(k,s)}-\frac{(t-j-1)}{2}-1}$$
$$\exp\left[-\frac{2\alpha_2^{(k,s)} + H_{j+1:t-1}}{2\sigma_{t-1}^2}\right]d\sigma_{t-1}^2$$
$$\propto \int (\sigma_{t-1}^2)^{-\alpha_1^{(k,s)}-\frac{(t-j)}{2}-1}\exp\left(-\frac{(y_t-\mu_*)^2 + (2\alpha_2^{(k,s)} + H_{j+1:t-1})\sigma_*^2}{2\sigma_{t-1}^2\sigma_*^2}\right)d\sigma_{t-1}^2$$
$$\propto \frac{\Gamma\left(\alpha_1^{(k,s)} + \frac{(t-j)}{2}\right)}{\left[\frac{(y_t-\mu_*)^2 + (2\alpha_2^{(k,s)} + H_{j+1:t-1})\sigma_*^2}{2\sigma_*^2}\right]^{\alpha_1^{(k,s)}+\frac{(t-j)}{2}}}$$
$$\propto \left[1 + \frac{1}{v}\frac{(y_t-\mu_*)^2v}{\sigma_*^2(2\alpha_2^{(k,s)} + H_{j+1:t-1})}\right]^{-\frac{1+v}{2}}$$

where $v = 2\alpha_1^{(k,s)} + t - j - 1$. Therefore

$$(y_t|\tau_t = j, s_t = s, k, y_{j+1:t-1}) \sim t_1\left(2\alpha_1^{(k,s)} + t - j\right.$$
$$\left. - 1, X_{t-j}\boldsymbol{\mu}_{j+1,t-1}, \frac{(2\alpha_2^{(k,s)} + H_{j+1:t-1})(1 + X_{t-j}\boldsymbol{\Sigma}_{j+1,t-1}X_{t-j}^T)}{2\alpha_1^{(k,s)} + t - j - 1}\right)$$

The proof for $j = t - 1$ is similar to the above derivation process and is neglected here.

REFERENCES

[1] 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.

[2] A. Heng, S. Zhang, A. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," *Mechanical Systems and Signal Processing,* vol. 23, pp. 724-739, 2009.

[3] M. S. Kan, A. C. Tan, and J. Mathew, "A review on prognostic techniques for non-stationary and non-linear rotating systems," *Mechanical Systems and Signal Processing,* vol. 62, pp. 1-20, 2015.

[4] M. Pecht and R. Jaai, "A prognostics and health management roadmap for information and electronics-rich systems," *Microelectronics Reliability,* vol. 50, pp. 317-323, 2010.

[5] 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.

[6] 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.

[7] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation–A review on the statistical data driven approaches," *European journal of operational research,* vol. 213, pp. 1-14, 2011.

[8] 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.

[9] C. J. Lu and W. O. Meeker, "Using degradation measures to estimate a time-to-failure distribution," *Technometrics,* vol. 35, pp. 161-174, 1993.

[10] W.-b. Wang, "A model to determine the optimal critical level and the monitoring intervals in condition-based maintenance," *International Journal of Production Research,* vol. 38, pp. 1425-1436, 2000.

[11] 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.

[12] N. Gebraeel, "Sensory-updated residual life distributions for components with exponential degradation patterns," *IEEE Transactions on Automation Science and Engineering,* vol. 3, pp. 382-393, 2006.

[13] 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.

[14] S. J. Bae and P. H. Kvam, "A nonlinear random-coefficients model for degradation testing," *Technometrics,* vol. 46, pp. 460-469, 2004.

[15] T. S. Ng, "An application of the EM algorithm to degradation modeling," *IEEE Transactions on Reliability,* vol. 57, pp. 2-13, 2008.

[16] 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.

[17] 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.

[18] C. Chiao and M. Hamada, "Using degradation data from an experiment to achieve robust reliability for light emitting diodes," *Quality and Reliability Engineering International,* vol. 12, pp. 89-94, 1996.

[19] S.-T. Tseng, M. Hamada, and C.-H. Chiao, "Using degradation data to improve fluorescent lamp reliability," *Journal of Quality Technology,* vol. 27, pp. 363-369, 1995.

[20] S. J. Bae and P. H. Kvam, "A change-point analysis for modeling incomplete burn-in for light displays," *IiE Transactions,* vol. 38, pp. 489-498, 2006.

[21] 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.

[22] J. Wu, Y. Chen, S. Zhou, and X. Li, "Online steady-state detection for process control using multiple change-point models and particle filters," *IEEE Transactions on Automation Science and Engineering,* vol. 13, pp. 688-700, 2016.

[23] N. Chopin, "Dynamic detection of change points in long time series," *Annals of the Institute of Statistical Mathematics,* vol. 59, pp. 349-366, 2007.

[24] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," *Handbook of nonlinear filtering,* vol. 12, p. 3, 2009.

[25] N. Whiteley, C. Andrieu, and A. Doucet, "Particle Markov Chain Monte Carlo for Multiple Change-point Problems," Technical Report 0911, Department of Mathematics, Bristol University2009.

[26] F. Caron, A. Doucet, and R. Gottardo, "On-line changepoint detection and parameter estimation with application to genomic data," *Statistics and Computing,* vol. 22, pp. 579-595, 2012.

[27] J. Wu, Y. Chen, and S. Zhou, "Online detection of steady-state operation using a multiple-change-point model and exact Bayesian inference," *IiE Transactions,* vol. 48, pp. 599-613, 2016.

[28] N. A. Heard and M. J. Turcotte, "Adaptive sequential Monte Carlo for multiple changepoint analysis," *Journal of Computational and Graphical Statistics,* pp. 1-26, 2016.

[29] P. Fearnhead, "Exact and efficient Bayesian inference for multiple changepoint problems," *Statistics and Computing,* vol. 16, pp. 203-213, 2006.

[30] J. Wu, Y. Liu, and S. Zhou, "Bayesian Hierarchical Linear Modeling of Profile Data With Applications to Quality Control of Nanomanufacturing," *IEEE Transactions on Automation Science and Engineering,* vol. 13, pp. 1355-1366, 2016.

[31] P. Bycott and J. Taylor, "A comparison of smoothing techniques for CD4 data measured with error in a time‐dependent Cox proportional hazards model," *Statistics in medicine,* vol. 17, pp. 2061-2077, 1998.

[32] M. S. Wulfsohn and A. A. Tsiatis, "A joint model for survival and longitudinal data measured with error," *Biometrics,* pp. 330-339, 1997.

[33] G. Schwarz, "Estimating the dimension of a model," *The annals of statistics,* vol. 6, pp. 461-464, 1978.

[34] J. Son, S. Zhou, C. Sankavaram, X. Du, and Y. Zhang, "Remaining useful life prediction based on noisy condition monitoring signals using constrained Kalman filter," *Reliability Engineering & System Safety,* vol. 152, pp. 38-50, 2016.

[35] A. Saxena, J. Celaya, E. Balaban, K. Goebel, B. Saha, S. Saha, and M. Schwabacher, "Metrics for evaluating performance of prognostic techniques," in *Prognostics and health management, 2008. phm 2008. international conference on,* 2008, pp. 1-17.

[36] N. Gebraeel, M. Lawley, R. Liu, and V. Parmeshwaran, "Residual life predictions from vibration-based degradation signals: a neural network approach," *IEEE Transactions on Industrial Electronics,* vol. 51, pp. 694-700, 2004.

[37] N. Z. Gebraeel and M. A. Lawley, "A neural network degradation model for computing and updating residual life distributions," *IEEE Transactions on Automation Science and Engineering,* vol. 5, pp. 154-163, 2008.

Yuxin Wen received the B.S. degree in Medical Informatics Engineering from Sichuan University, Sichuan, China in 2011, the M.S. degree in Biomedical Engineering from Zhejiang University, Zhejiang, China in 2014. Currently, she is pursuing the Ph.D. degree in Electrical and Computer Engineering at the University of Texas at El Paso (UTEP), TX, USA. Her research interests are focused on statistical modeling, prognostics and reliability analysis.
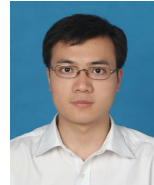
Jianguo Wu is an Assistant Professor in the Department of Industrial, Manufacturing and Systems Engineering (IMSE) at UTEP, TX, USA. He received the B.S. degree in Mechanical Engineering from Tsinghua University, Beijing, China in 2009, the M.S. degree in Mechanical Engineering from Purdue University, West Lafayette, IN, USA in 2011, and M.S. degree in Statistics in 2014 and Ph.D. degree in Industrial and Systems Engineering in 2015, both from University of Wisconsin-Madison, WI, USA. His research interests are focused on statistical modeling, monitoring and analysis of complex processes/systems for quality control and productivity improvement through integrated application of metrology, engineering domain knowledge and data analytics. He is a member of INFORMS, IISE, and SME.

Qiang Zhou is an Assistant Professor in the Department of Systems and Industrial Engineering, The University of Arizona. He received a B.S. degree in Automotive Engineering (2005) and a M.S. degree in Mechanical Engineering (2007) from Tsinghua University, a M.S. degree in Statistics (2010) and a Ph.D. degree in Industrial Engineering (2011) at the University of Wisconsin-Madison. His research interests include modeling, monitoring and analysis of complex engineering systems for the purpose of quality control and productivity improvement. Dr. Zhou is a member of IEEE, INFORMS, IIE, and ASQ.

Tzu-Liang (Bill) Tseng is a Professor and Chair of Department of IMSE at UTEP. He received his M.S. degree in Industrial Engineering from the University of Wisconsin-Madison in 1993 and 1995 respectively and Ph.D. in Industrial Engineering from the University of Iowa in 1999. His research area cover quality assurance in additive manufacturing, industrial data analytics and cyber based decision support systems. Dr. Tseng is currently serving as an editor of Journal of CSI and editor boards of JDMMM and AJIBM. He is currently a Senior Member of IISE, SME and the Program Chair of Manufacturing Division of ASEE.