

Chapman University

Chapman University Digital Commons

Engineering Faculty Articles and Research

Fowler School of Engineering

6-15-2017

Multiple-Phase Modeling of Degradation Signal for Condition Monitoring and Remaining Useful Life Prediction

Yuxin Wen

Jianguo Wu

Yuan Yuan

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](#)

Multiple-Phase Modeling of Degradation Signal for Condition Monitoring and Remaining Useful Life Prediction

Comments

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in *IEEE Transactions on Reliability*, volume 66, issue 3, in 2017 following peer review. The definitive publisher-authenticated version is available online at <https://doi.org/10.1109/TR.2017.2710319>.

Copyright

© 2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Multiple Phase Modeling of Degradation Signal for Condition Monitoring and Remaining Useful Life Prediction

Yuxin Wen, Jianguo Wu, Yuan Yuan

Abstract—Remaining useful life prediction plays an important role in ensuring the safety, availability and efficiency of various engineering systems. In this paper, we propose a flexible Bayesian multiple phase modeling approach to characterize degradation signals for prognosis. The priors are specified with a novel stochastic process and the multiple phase model is formulated to a novel state-space model to facilitate online monitoring and prediction. A particle filtering algorithm with stratified sampling and partial Gibbs resample-move strategy is developed for online model updating and residual life prediction. The advantages of the proposed method are demonstrated through extensive numerical studies and real case studies.

Index Terms—Condition monitoring, remaining useful life prediction, multiple change-point model, particle filters, prognostics and health management

ABBREVIATIONS AND ACRONYMS

RUL	Remaining useful life
CM	Condition monitoring
BIC	Bayesian information criterion
<i>i.i.d.</i>	Independent and identically distributed
MLE	Maximum likelihood estimate
PF	Particle filtering
SIS	Sequential importance sampling
SPF	Stratified particle filtering
RMSD	Root mean square deviation

NOTATION

k_i	Number of change-points for unit i
\mathbf{Y}_i	Vector of observations for unit i
T_i	Number of observations for unit i
c_{ik_i}	Location of the k_i -th change-point for unit i
a_{ik}	Intercept of the k -th line segment for unit i
b_{ik}	Slope of the k -th line segment for unit i

σ_{ik}	Standard deviation of the k -th line segment for unit i
$y_{1:t}$	Observations of a working unit by the current time index t
\mathbf{x}_t	State vector at the current time t
s_t	Phase or stage index at time t
$\pi(\cdot)$	Prior distribution
τ_t	Latest change-point at time t
$\delta^{(s)}$	Duration of the s -th segment
$\boldsymbol{\psi}$	Hyper-parameters of all priors
$\boldsymbol{\theta}^{(s)}$	Model parameters of the s -th segment
$p_{t+1}(\mathbf{x}_t)$	Prior transition probability of state vector \mathbf{x}_t at time $t + 1$
$\boldsymbol{\beta}^{(s)}$	Vector of coefficients including slope and intercept
$\hat{\boldsymbol{\beta}}^{(s)}$	Estimated coefficients including slope and intercept
$\hat{\sigma}^{2(s)}$	Estimated measurement noise
$\boldsymbol{\mu}_t^{(k,s)}$	Mean of normal distribution of the s -th segment
$\boldsymbol{\Sigma}_t^{(k,s)}$	Covariance of normal distribution of the s -th segment
$\alpha_1^{(k,s)}$	Shape parameter of inverse Gamma distribution of the s -th segment
$\alpha_2^{(k,s)}$	Scale parameter of inverse Gamma distribution of the s -th segment
l_k	Positive lower bound of the slope of the last phase for k -change-point case
$f(\cdot \cdot)$	Prior transition probability density function of state-space model
$g(\cdot \cdot)$	Density function of the observation model
N	Number of particles
$W_t^{(i)}$	Normalized weight for each particle i
$W_t^{(g)}$	Group weight coefficient for each group g
R_t	Remaining useful life at the current time t
Γ	Failure threshold
$\Phi(\cdot)$	Cumulative distribution function of Gaussian distribution

I. INTRODUCTION

THE remaining useful life (RUL) refers to the time remaining before a failure occurs at a particular time of operation [1]. RUL prediction plays a critical role in support of executing preventive actions [2], replacement strategies [3] and health management [4], as well as maximizing manufactured products availability [5]. Condition monitoring (CM) signals, also referred as degradation signals, are often in situ collected during system operations. They are directly related to the health condition of the system and have been widely used for condition monitoring and RUL prediction in the past few decades [6]. According to Zio *et al.*[7] and Dragomir *et al.*[8],

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, and ECE Department (affiliated), University of Texas at El Paso, Texas 79968, USA (e-mail: jwu2@utep.edu).

Y. Yuan is with the IBM Research, Singapore (e-mail: polarisyy@gmail.com)

the existing scientific literature on data-driven modeling and prediction can be classified into two categories, the statistical and artificial intelligence approaches. The statistical approaches include regression-based, Wiener process, and Markovian based models, while the artificial intelligence approaches include neural networks, decision tree, support vector machines, fuzzy system, etc. Due to the excellent statistical properties and interpretation characteristics, statistical-based methods have become more and more popular. A comprehensive review of statistical approaches can be found in Si *et al.* [9].

The commonly applied statistical approach is to fit CM signals using parametric regression models to describe and predict how the currently available CM signal evolves [7]. In these methods, the functional forms of the models are often linear, polynomial, exponential, or combination of them [10]. However, these parametric models are often too rigid and not adequate or flexible enough to model the real CM signals in the whole time period. For example, Son *et al.* [11] showed that the internal resistance degradation signal of vehicle batteries changes its functional form and evolves more rapidly after the system has degraded down to a certain level before failure occurs. Bae and Kvam [12] also demonstrated that the degradation path of vacuum fluorescent displays is not monotonic, showing obviously two phases or even three phases. This phenomenon has also been observed in many other CM signals, such as fatigue-crack-growth data [13], the thrust-force data of drill-bits [14], and vibrational signals of rotational bearings [15]. Some researchers chose to delete early degradation measurements at the first stage, under the assumption that the failure will not occur at the early stage, and then apply the parametric models to the second phase data for better model fitting and prediction [16, 17]. However, the truncated measurements may contain valuable information about the degradation process or the prediction may need to be made at the early stage.

To avoid measurement truncation, some researchers proposed to add a change-point to divide CM signals into two phases and fit each phase with different models. Bae *et al.* [18] found that the prediction accuracy can be improved substantially by the addition of a change-point for modeling incomplete burn-in data of light displays. Li and Nilkitsaranont [19] employed a combination of a linear model in the first phase and a quadratic model in the second phase to estimate the remaining useful life of gas turbine engines, and used “compatibility check” to determine the transition point from one model to another. Gebraeel *et al.* [6] developed an exponential (i.e., log-linear) degradation model with a pre-set location of a change-point to illustrate the updating process of rolling element bearings. Later Chen and Tsui [15] revisited Gebraeel’s work [6], and applied a two-phase regression model with one change-point at unknown location to characterize both phases of the bearing degradation signals.

All these aforementioned methods assume a two-phase pattern on CM signals. In many situations, however, the degradation path may have three or even more phases during the whole life cycle. It would be difficult to select proper

functional forms to characterize the degradation behavior with no change-point or only one change-point. To fill such gap, this paper endeavors to develop a multiple-phase modeling approach for condition monitoring and RUL prediction. Here it is necessary to discuss more about the term “phase”. In the existing literature, the definition of “phase” can be classified into two categories: (1) it is commonly referred to as different operational conditions or states, such as the working state and storage state for missiles [20], multiple consecutive phases of operations required to finish the service for phased-mission systems (PMS) [21]; (2) it is also frequently used to denote health conditions with different characteristics (e.g., normal working stage and irreversible degradation stage with defects occurred for bearings [6, 22]), or different patterns shown on CM signals, which may not have specific physical meanings [11, 15]. In our approach, “multiple-phase” is more related to the second category, though it can be easily applied to the first case as long as the degradation signals exist multiple patterns. The main difference between our work and the existing multiple-phase approaches by Si *et al.* [20, 21] lies in the motivation and methodologies. In Si *et al.*’s work, as mentioned earlier, the multiple phases are used to model different operational states or stages, e.g., take-off, ascent, cruise, approach and landing phases of the on-board systems for the aided-guide of aircraft. Therefore, in these methods, the number of phases are fixed, the phase index and the starting point of each phase before the current time can be exactly observed. However, in our work, the purpose is to provide a flexible multiple change-point based approach to model highly nonlinear degradation signals where the existing functional forms are inadequate or not applicable. Therefore, each phase may not have any physical meanings, the number of phases may not be fixed, and both the phase index and its starting time are random variables across all the life cycle and need to be estimated.

To characterize the population-level trend as well as the individual heterogeneity, mixed-effects or random-effects models are most commonly selected in off-line modeling of historic CM signals. When predicting the RUL of a new unit, the Bayesian approach is naturally selected for online model updating, prediction and uncertainty quantification, where the fitted parameters in the offline stage are used as priors [15]. However, there are several challenges on how to effectively apply the multiple change-point model for condition monitoring and RUL prediction under the Bayesian framework. Due to significantly increased dimensionality and complexity, it is difficult to specify reasonable priors (e.g., phase durations or number of change-points, model parameters of each phase) in the offline modeling of historical CM signals. In addition, in the online model updating stage, the posterior distributions of model parameters need to be updated sequentially. However, the multiple change-point model is highly nonlinear and the conventional Kalman filtering techniques [23], which are commonly used for linear models, are not applicable. Besides, the RUL prediction given the posterior of the current model parameters is still very challenging due to the uncertainty of future change-points and model parameters. To address these

challenges, we propose a series of approaches in both off-line modeling and online model updating and RUL prediction. In the off-line modeling, a novel stochastic process is proposed to specify and estimate priors in the off-line modeling. In the online stage, the multiple change-point model is formulated as a non-standard state-space model and a novel stratified particle filtering algorithm is developed for online model updating and RUL prediction. The contribution of this paper lies in the following three-fold: (1) we innovatively apply the multiple change-point model to degradation signals to improve modeling and prognostics, which is fundamentally different from the existing multiple-phase modeling approaches in terms of motivation, methodology and applications; (2) a full Bayesian framework is proposed for the multiple change-point model through a novel stochastic process; and (3) an efficient stratified particle filtering algorithm with partial Gibbs sampling strategy is developed for model updating and RUL prediction.

The rest of the paper is organized as follows. In Section 2, a multiple change-point model for the CM signals is presented. The prior parameters specification and state-space representation for the multiple change-point model are given in Section 3. Section 4 presents the technical details on how to sequentially update the posterior distributions of the phase index, latest change-point, and model parameters of the current phase, and how to predict the RUL using the particle filtering algorithm. 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 MODEL FOR DEGRADATION SIGNAL

In this paper, a piece-wise linear functional form is proposed to model CM signals. The piecewise linear model with a proper number of change-points at proper locations is flexible enough to capture the non-linear and multiple-phase characteristics of various kinds of degradation signals in application. It could avoid the nontrivial selection of appropriate functional forms to model the CM signals. Besides, it makes more sense physically to define it as a phase when CM signal is degrading with a constant rate.

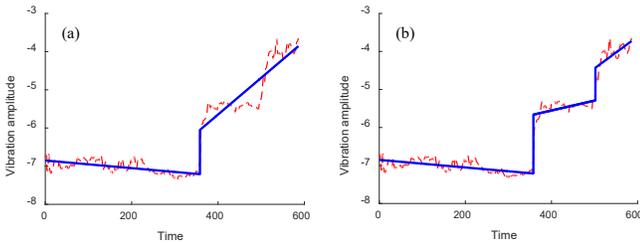


Figure 1. Piecewise linear model for bearing signal: (a) one change-point; (b) two change-points.

Here we use the bearing vibration signals [6, 15] to demonstrate the superiority of the proposed method. As shown in Figure 1 (a), the second phase of the bearing data is poorly fitted if only one change-point is introduced. It can be seen that

there are two abrupt changes on this bearing CM signal before it hits the pre-specified failure threshold. From Figure 1 (b) we can clearly see that a three-phase model is much more accurate. Note that the model accuracy may significantly influence the accuracy of the RUL prediction. For example, if only one change-point is adopted, then at a certain time between 400 and 500, the fitted line of the second phase would be very flat, which will cause the predicted RUL to be significantly larger than the actual value.

In some applications, all units may have the same degrading behavior, i.e., they will experience the same number of degrading phases before failure. However, it is also common in practice that the units are heterogeneous, and the number of change-points required across all units could be different. Therefore, we assume the number of phases is random to make our model more flexible. Denote k_i as the number of change-points for unit i with T_i observations $\mathbf{Y}_i = \{y_{i1}, y_{i2}, \dots, y_{iT_i}\}$. The change-points are denoted as $c_{i1}, c_{i2}, \dots, c_{ik_i}$ which satisfy $c_{i0} = 0 < c_{i1} < c_{i2} < \dots < c_{ik_i} < c_{ik_i+1} = T_i$. As there are k_i change-points, there are in total $k_i + 1$ line segments. Mathematically, the piece-wise linear model can be expressed as

$$y_{ij} = \begin{cases} a_{i1} + b_{i1}t_{ij} + \sigma_{i1}\varepsilon_{ij}, & 0 < t_{ij} \leq c_{i1} \\ a_{i2} + b_{i2}t_{ij} + \sigma_{i2}\varepsilon_{ij}, & c_{i1} < t_{ij} \leq c_{i2} \\ \dots & \dots \\ a_{ik_i} + b_{ik_i}t_{ij} + \sigma_{ik_i}\varepsilon_{ij}, & c_{ik_i-1} < t_{ij} \leq c_{ik_i} \\ a_{ik_i+1} + b_{ik_i+1}t_{ij} + \sigma_{ik_i+1}\varepsilon_{ij}, & c_{ik_i} < t_{ij} \leq T_i \end{cases} \quad (1)$$

where ε_{ij} follows *i.i.d.* standard normal distribution, a_{ik}, b_{ik} and σ_{ik} are the intercept, slope and standard deviation of the k -th line segment, respectively. Without loss of generality, we assume that all the sampling intervals equal to 1 for all units, i.e., $t_{ij} = j$ for the sake of simplicity.

Bayesian approach is commonly employed to integrate historical data with newly observed CM signal of a working unit for sequential model updating and RUL prediction. The historical data provide prior information on the number of change-points, locations of change-points, and possible values of model parameters of each line segment. Based on the prior information and observed CM signal of a working unit up to the current time, the posterior distribution of the model parameters of the current line segment and future observations can be updated. Denote $y_{1:t}$ as the observations of a working unit by the current time index t , and denote $\mathbf{x}_t = (a_t, b_t, \sigma_t^2, \tau_t, s_t, k)$ where a_t, b_t, σ_t^2 are the parameters of the current line segment, τ_t is the latest change-point, i.e., the starting time of the current line segment, s_t is the phase or stage index, and k is the total number of phases in the whole life cycle. In the model updating, the posterior of the current line segment can be expressed as

$$P(\mathbf{x}_t | y_{1:t}) \propto \pi(\mathbf{x}_t) P(y_{1:t} | \mathbf{x}_t) \quad (2)$$

where $\pi(\cdot)$ is the prior distribution obtained in the offline modeling of historical data. To predict the RUL, the posterior distribution of the future degradations $y_{t^*}, \forall t^* > t$ needs to be calculated based on the updated model

$$P(y_{t^*}|y_{1:t}) = \sum_{k, s_t, \tau_t} \int P(y_{t^*}|\mathbf{x}_t)P(\mathbf{x}_t|y_{1:t})d(a_t, b_t, \sigma_t^2) \quad (3)$$

where

$$P(y_{t^*}|\mathbf{x}_t) = \sum_{s_t^*, \tau_t^*} \int P(y_{t^*}|\mathbf{x}_{t^*})\pi(\mathbf{x}_{t^*}|\mathbf{x}_t)d(a_{t^*}, b_{t^*}, \sigma_{t^*}^2) \quad (4)$$

Although the above three equations have simple formulation, they are generally intractable due to high dimensionality and high nonlinearity caused by the unknown change-points. To address this challenge, we reformulate the multiple change-point model to a nonstandard state-space model and use particle filtering techniques to approximate these posteriors. Section 3 will introduce the specification and calculation of priors based on the historical data, and the state-space representation of the multiple change-point model. Section 4 will give the technical details of the developed particle filtering algorithm.

III. PRIOR SPECIFICATION AND STATE-SPACE REPRESENTATION

Denote a multiple change-point model as $\mathcal{M} = (k, \{\delta^{(s)}\}_{s=1}^{k+1}, \{\theta^{(s)}\}_{s=1}^{k+1})$ where k is the number of change-points, $\delta^{(s)}$ is the duration of the s -th segment, i.e., $\delta^{(s)} = c_s - c_{s-1}$, and $\theta^{(s)}$ is the model parameters of the s -th segment, e.g., $\theta^{(s)} = (a^{(s)}, b^{(s)}, \sigma^{2(s)})$ in this paper. In the Bayesian formulation of multiple change-point models with a *fixed* number of observations, the priors for the number of change-point k , the segment durations $\{\delta^{(s)}, s = 1, \dots, k+1\}$ and the segment parameters $\{\theta^{(s)}, s = 1, \dots, k+1\}$ can be specified easily. For the change-points, a joint prior could be placed, i.e., $\pi(k, \{\delta^{(s)}\}_{s=1}^{k+1}) = \pi(k)\pi(\{\delta^{(s)}\}_{s=1}^{k+1}|k)$. More commonly, a marked renewal process could be assumed to simultaneously model the priors for the number of change-points and their occurrence intervals or equivalently their locations [24]. For example, a Poisson process could be used to model the occurrence of change-points, where the successive intervals $\delta^{(s)}$ follow an *i.i.d.* exponential distribution, and the last interval satisfies $\delta^{(k+1)} \geq T - c_k$. In such case, the prior density can be derived as

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

where $f(\cdot|\lambda)$ is the probability density function of an exponential distribution, λ is the Poisson rate and T is the total number of observations. In a Bernoulli process, each time step has the probability p to be a change-point and the interval times

follow an *i.i.d.* geometric distribution [25-27]. The joint density is simply

$$\pi(k, \{\delta^{(s)}\}_{s=1}^{k+1}) = p^k(1-p)^{T-1-k}$$

where p is the parameter for the Bernoulli distribution. For the changing parameters $(\theta^{(s)}, s = 1, \dots, k)$, *i.i.d.* Gaussian distribution is often assigned.

The aforementioned renewal process is often applied in the segmentation of time series data of a known and fixed length and the priors specified are often non-informative, i.e., the phase duration follows the same distribution. However, considering the phase heterogeneity of the CM signals, the prior distributions for the phase durations should be different to make the prior more informative for RUL prediction. Also, for a working unit, the number of observations to be collected before it fails is unknown. If a renewal process is applied to model the priors, an unlimited number of change-points beyond the current time has to be considered, which is unrealistic for informative prior specifications and RUL prediction. To solve this problem, we first place a prior distribution on the number of change-point k . Conditioning on k , the distribution of the phase interval lengths are modelled by a stochastic process where the first k interval lengths $\{\delta^{(s)}\}_{s=1}^k$ follow independent and non-identical distributions and the $k+1$ model parameters $\{\theta^{(s)}\}_{s=1}^{k+1}$ follow independent and non-identical distributions. Then the prior could be factorized as

$$\pi(\mathcal{M}) = \pi(k) \prod_{s=1}^k \pi(\delta^{(s)}|k) \prod_{s=1}^{k+1} \pi(\theta^{(s)}|k) \quad (5)$$

More specifically, we put a categorical distribution or generalized Bernoulli distribution on k , with $\pi(k = k_i) = p_i$ and $\sum_i p_i = 1$. For simplicity, we assume the phase durations follow normal distributions, $\delta^{(s)}|k \sim \mathcal{N}(\delta_0^{(k,s)}, \sigma_0^{2(k,s)})$. For the changing parameters, the commonly used normal and inverse Gamma are specified, i.e., $\beta^{(s)}|k = (a^{(s)}, b^{(s)}|k) \sim \mathcal{N}(\mu_0^{(k,s)}, \Sigma_0^{(k,s)})$ and $\sigma^{2(s)}|k \sim IG(\alpha_1^{(k,s)}, \alpha_2^{(k,s)})$. Since the CM signal often increases rapidly when it is approaching the failure threshold in the last phase, we assume a truncated normal prior for the last segment to make the prior more informative: $\beta^{(k+1)}|k \sim TN(\mu_0^{(k,k+1)}, \Sigma_0^{(k,k+1)}|b^{(k+1)} > l_k)$ where l_k is a positive lower bound of the slope for k -change-point case. Note that here we assume the model parameters are independent across different phases to reduce both the computational complexity and the required number of historical CM signals.

To specify informative priors, the hyper-parameters of all these priors, denoted as ψ , need to be estimated based on the historical data. One common way is to estimate ψ by maximizing the marginal likelihood [15] of I historical CM signals

$$\hat{\boldsymbol{\psi}} = \arg \max_{\boldsymbol{\psi}} \prod_{i=1}^I \int P(Y_i | \mathcal{M}_i) \pi(\mathcal{M}_i | \boldsymbol{\psi}) d\mathcal{M}_i$$

Unfortunately, the marginal likelihood is very complex and not tractable. An alternative approach is a two-stage process where the model parameters $\hat{\mathcal{M}}_i$ of each unit i are first obtained through the maximum likelihood estimates (MLE) and then the hyper-parameters are estimated through the MLE by treating these estimated models $\{\hat{\mathcal{M}}_i, i = 1, \dots, I\}$ as observations. In our case, however, MLE cannot be directly applied to each CM signal since increasing k will also increase the fitting accuracy, and thus result in over-fitting. To address this issue, we propose to use Bayesian information criterion (BIC) [28] for model selection and parameter estimation of each CM signal

$$\hat{\mathcal{M}} = \arg \min_{\mathcal{M}} (-2l(\mathcal{M} | \mathbf{Y}) + n \log T) \quad (6)$$

where $n = 4k + 3$ is the number of model parameters to estimate, $l(\mathcal{M} | \mathbf{Y})$ is the log-likelihood given as

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

where

$$X_{c_{s-1}+1:c_s} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \tau_{c-1} + 1 & \tau_{c-1} + 2 & \dots & c_s \end{bmatrix}^T \quad (8)$$

Given the number of change-points and their locations $\{k, c_1, \dots, c_k\}$, the parameters $\{\boldsymbol{\beta}^{(s)}, \sigma^{2(s)}, s = 1, \dots, k + 1\}$ that minimize the BIC are just the MLE of the Gaussian linear models of each phase

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{(s)} &= (X_{c_{s-1}+1:c_s}^T X_{c_{s-1}+1:c_s})^{-1} X_{c_{s-1}+1:c_s}^T y_{c_{s-1}+1:c_s}^T \\ \hat{\sigma}^{2(s)} &= \|y_{c_{s-1}+1:c_s}^T - X_{c_{s-1}+1:c_s} \hat{\boldsymbol{\beta}}^{(s)}\|^2 / \delta^{(s)} \end{aligned} \quad (9)$$

If k and T are small, it is possible to try all combinations $\{k, c_1, \dots, c_k\}$ to determine the optimal model. However, this method is not realistic for large k and T due to the exponentially increased computational cost. Instead we could use the PELT method [29], which is computationally efficient with a computational cost that is linear with T .

Based on the above prior specification, the multiple change-point model could be formulated to a non-standard state-space model with state vector $\mathbf{x}_t = (\boldsymbol{\theta}^{(s_t)}, \tau_t, s_t, k)$ and prior state transition process

$$\mathbf{x}_{t+1} = \begin{cases} \mathbf{x}_t, & p = 1 - p_{t+1}(\mathbf{x}_t) \text{ if } s_t < k + 1 \\ \mathbf{x}_t, & p = 1 \text{ if } s_t = k + 1 \\ (\boldsymbol{\theta}^{(s_t+1)}, t, s_t + 1, k) & p = p_{t+1}(\mathbf{x}_t) \text{ if } s_t < k + 1 \end{cases} \quad (10)$$

Here $p_{t+1}(\mathbf{x}_t)$ is the transition probability of the stochastic process expressed by

$$p_{t+1}(\mathbf{x}_t) = \frac{p(\delta^{(s_t)} \leq L + 1 | \delta^{(s_t)} \geq L) \Phi(L + 1 | \delta_0^{(k, s_t)}, \sigma_0^{2(k, s_t)}) - \Phi(L | \delta_0^{(k, s_t)}, \sigma_0^{2(k, s_t)})}{1 - \Phi(L | \delta_0^{(k, s_t)}, \sigma_0^{2(k, s_t)})} \quad (11)$$

where $L = t - \tau_t$ and $\Phi(\cdot)$ is the Gaussian cumulative distribution function. Note that when a hidden state is continuous-valued, the term *state-space model* is often used instead of *hidden Markov model*. Here we refer to our model as a *non-standard state-space model* in that its state vector contains both discrete and continuous-valued components, and the state \mathbf{x}_{t+1} is not linearly correlated with \mathbf{x}_t , which is different from standard state-space model.

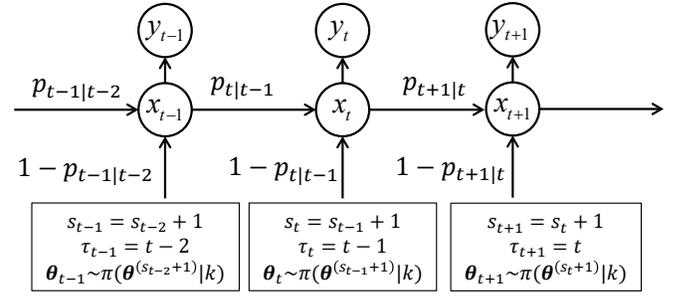


Figure 2. Illustration of the formulated state-space model.

The formulated state-space model is illustrated in Figure 2, where the transition probability from \mathbf{x}_t to \mathbf{x}_{t+1} can be expressed as

$$p_{t+1|t} = [1 - p_{t+1}(\mathbf{x}_t)]^{1(s_t < k+1)} \quad (12)$$

If there are no change-points, the formulated state-space model is a special linear state-space model with a constant state, which can be easily inferred using Kalman filters. However, due to the existence of unknown change-points, the formulated state-space is highly nonlinear, which makes the inference very challenging. The particle filtering techniques are particularly effective for nonlinear state-space models and have been widely applied in the prognosis area. Generally, the applications in the prognosis area can be classified into three categories based on the underlying state-space model: (1) nonlinear state transition model, linear observation model [30]; (2) linear state transition model, nonlinear observation model [31, 32] and (3) nonlinear state transition model and nonlinear observation model [33, 34]. The formulated state-space model in this paper falls into the third category. However, it is fundamentally different from the existing ones due to its special characteristics, i.e., high dimensionality, containing both discrete and continuous states, some states being constant (linear transition) across all life cycle while some states being constant between two successive change-points but changing once a new change-point occurs (nonlinear transition). To our best knowledge, none of the existing algorithms work well on

our model. In the following section, a novel stratified particle filtering algorithm with partial Gibbs sampling strategy will be developed for sequential model updating and RUL prediction.

IV. PARTICLE FILTERING ALGORITHM FOR ONLINE MODEL UPDATING AND RUL PREDICTION

Particle filters (PF) are effective sequential Monte Carlo (SMC) methods to solve the filtering problems. It is particularly useful for sequential Bayesian inference of linear/nonlinear Gaussian/non-Gaussian state-space models [35]. In this section, a customized particle filtering algorithm is developed for sequential model estimation and RUL prediction of a working unit. For the sake of completeness, the basic theory of PF algorithm is first presented.

A. Review of Particle Filtering Algorithm

The basic idea of the PF technique is the sequential importance sampling (SIS). Consider a state-space model described as

$$\mathbf{x}_1 \sim f(\mathbf{x}), \mathbf{x}_t | \mathbf{x}_{t-1} \sim f(\mathbf{x}_t | \mathbf{x}_{t-1}), y_t | \mathbf{x}_t \sim g(y_t | \mathbf{x}_t) \quad (13)$$

where $f(\cdot)$ is the prior for the first state and $f(\cdot | \cdot)$ is the prior state transition probability density associated with state changing from \mathbf{x}_{t-1} to \mathbf{x}_t , and $g(\cdot | \cdot)$ is the density function of y_t conditioning on \mathbf{x}_t . The observations $y_{1:T}$ are assumed to be conditionally independent given $\mathbf{x}_{1:T}$. According to Bayes' theorem, the posterior density satisfies the following recursion

$$p(\mathbf{x}_{1:t} | y_{1:t}) = p(\mathbf{x}_{1:t-1} | y_{1:t-1}) \frac{f(\mathbf{x}_t | \mathbf{x}_{t-1}) g(y_t | \mathbf{x}_t)}{p(y_t | y_{1:t-1})} \quad (14)$$

where

$$p(y_t | y_{1:t-1}) = \int p(\mathbf{x}_{t-1} | y_{1:t-1}) f(\mathbf{x}_t | \mathbf{x}_{t-1}) g(y_t | \mathbf{x}_t) d\mathbf{x}_{t-1} \quad (15)$$

In the filtering problem, $p(\mathbf{x}_t | y_{1:t})$ is of interest and can be obtained by integrating out $\mathbf{x}_{1:t-1}$ or directly based on Bayes' theorem

$$p(\mathbf{x}_t | y_{1:t}) = \frac{p(\mathbf{x}_t | y_{1:t-1}) g(y_t | \mathbf{x}_t)}{p(y_t | y_{1:t-1})} \quad (16)$$

where

$$p(\mathbf{x}_t | y_{1:t-1}) = \int p(\mathbf{x}_{t-1} | y_{1:t-1}) f(\mathbf{x}_t | \mathbf{x}_{t-1}) d\mathbf{x}_{t-1} \quad (17)$$

Equation (17) is known as the prediction step and (16) is called as the updating step. However, Equation (15) and (17) are often intractable analytically, and SIS is often used for posterior approximation. If we select an important distribution that can be sequentially sampled with the following structure

$$q_t(\mathbf{x}_{1:t}) = q_1(\mathbf{x}_1) \prod_{i=2}^t q_i(\mathbf{x}_i | \mathbf{x}_{1:i-1}) \quad (18)$$

then the unnormalized weight function can be expressed by

$$\begin{aligned} \omega_t(\mathbf{x}_{1:t}) &= \frac{p(\mathbf{x}_{1:t}, y_{1:t})}{q_t(\mathbf{x}_{1:t})} \\ &= \frac{f(\mathbf{x}_1) \prod_{i=2}^t f(\mathbf{x}_i | \mathbf{x}_{i-1}) \prod_{i=1}^t g(y_i | \mathbf{x}_i)}{q_1(\mathbf{x}_1) \prod_{i=2}^t q_i(\mathbf{x}_i | \mathbf{x}_{1:i-1})} \quad (19) \\ &= \omega_1(\mathbf{x}_1) \prod_{i=2}^t w_i \end{aligned}$$

where

$$\begin{aligned} \omega_1(\mathbf{x}_1) &= f(\mathbf{x}_1) g(y_1 | \mathbf{x}_1) / q_1(\mathbf{x}_1) \\ w_i &= g(y_i | \mathbf{x}_i) f(\mathbf{x}_i | \mathbf{x}_{i-1}) / q_i(\mathbf{x}_i | \mathbf{x}_{1:i-1}) \end{aligned} \quad (20)$$

Equation (19) shows that the weight function can be calculated recursively, so that the posterior could be sequentially updated once a new observation is measured. The expectation of any function $\varphi(\mathbf{x}_{1:t})$ with respect to the posterior $p(\mathbf{x}_{1:t} | y_{1:t})$ can be estimated by

$$E(\varphi(\mathbf{x}_{1:t}) | y_{1:t}) \approx \sum_{i=1}^N W_t^{(i)} \varphi(\mathbf{x}_{1:t}^{(i)}) \quad (21)$$

where $W_t^{(i)}$ is the normalized weight. In the PF algorithm, a resampling step based their updated weights is often added to obtain equally weighted particles which are approximately distributed as $p(\mathbf{x}_{1:t} | y_{1:t})$. It is a ‘‘Darwinian’’ procedure that can remove particles with low weights and carry on particles with high weights. The generic particle filtering algorithm with a resampling step is given in Algorithm 1 as follows.

Algorithm 1. Generic Particle Filtering Algorithm

At time $t = 1$,

1. Sample $\mathbf{x}_1^{(i)} \sim q_1(\mathbf{x}_1)$ for $i = 1, \dots, N$
2. Compute weights $\omega_1(\mathbf{x}_1^{(i)})$ and normalized weights $W_1^{(i)} = \omega_1(\mathbf{x}_1^{(i)}) / \sum_{i=1}^N \omega_1(\mathbf{x}_1^{(i)})$.
3. Resample $\{W_1^{(i)}, \mathbf{x}_1^{(i)}\}$ according to their weights to obtain N equally weighted particles $\{\frac{1}{N}, \bar{\mathbf{x}}_1^{(i)}\}$ and set $\{W_1^{(i)}, \mathbf{x}_1^{(i)}\} \leftarrow \{\frac{1}{N}, \bar{\mathbf{x}}_1^{(i)}\}$.

At time $t \geq 2$:

1. Sample $\mathbf{x}_t^{(i)} \sim q_t(\mathbf{x}_t | \mathbf{x}_{1:t-1}^{(i)})$, set $\mathbf{x}_{1:t}^{(i)} \leftarrow (\mathbf{x}_{1:t-1}^{(i)}, \mathbf{x}_t^{(i)})$ for $i = 1, \dots, N$
 2. Compute $w_t(\mathbf{x}_{1:t}^{(i)})$ and normalized weights $W_t^{(i)}$
 3. Resample $\{W_t^{(i)}, \mathbf{x}_{1:t}^{(i)}\}$ to obtain N equally weighted particles $\{\frac{1}{N}, \bar{\mathbf{x}}_{1:t}^{(i)}\}$ and set $\{W_t^{(i)}, \mathbf{x}_{1:t}^{(i)}\} \leftarrow \{\frac{1}{N}, \bar{\mathbf{x}}_{1:t}^{(i)}\}$
-

B. Stratified Particle Filtering Algorithm for Model Updating

In the development of PF algorithm, the importance function needs to be specified. The optimal importance function should be the one that minimizes the variances of the importance weight of sampled particles [35]. It can reduce the particle

degeneracy issue, i.e., the weights concentrate on only a few particles and most particles have negligible weights. However, the optimal importance function is often not obtainable in practice. Instead, we propose to use the prior transition density as the importance density function

$$q_t(\mathbf{x}_t|\mathbf{x}_{1:t-1}) = f(\mathbf{x}_t|\mathbf{x}_{t-1}) \quad (22)$$

Selecting the prior transition density is the most common and convenient way in practical applications. According to Equation (20), with the resampling step implemented, the weight is simply

$$\omega_t(\mathbf{x}_{1:t}) = g(y_t|\mathbf{x}_t) \quad (23)$$

at each step, which greatly simplifies the computation.

Another important issue commonly faced in PF techniques is the particle impoverishment problem [25], where the number of unique particles or unique components of particles becomes less and less along iterations due to the resampling step. In our state-space model, each line segment between successive change-points is a special linear state-space model with a constant state, which makes the particle impoverishment problem even worse. Besides, the dimension of the state vector is relatively high, which may require a significantly large number of particles to guarantee the approximation accuracy, and thus result in high computational cost. However, for online model updating and RUL prediction, a low computational cost is often critically important. In this paper, we propose a Gibbs resample-move step to address both these issues. The resample-move strategy was first proposed by Gilks [36], where a “move” step is added after the resampling step to generate new particles through MCMC kernels with the posterior distribution as the invariant distributions. It can not only diversify the particles to reduce the particle impoverishment issue, but can also generate more particles with significant weights, thus reducing the particle degeneracy issue and reducing the required number of particles. In our algorithm, we propose a one-step partial Gibbs sampler to “move” the three continuous components $(\boldsymbol{\beta}_t, \sigma_t^2)$ through their conditional posterior distributions. The conditional posterior distributions are obtained based Lemma 1 as follows (the proof is included in the Appendix).

Lemma 1. Suppose $\boldsymbol{\beta}^{(s)}|k = (a^{(s)}, b^{(s)}|k)' \sim N(\boldsymbol{\mu}_0^{(k,s)}, \boldsymbol{\Sigma}_0^{(k,s)})$ for $s = 1, \dots, k$, $\boldsymbol{\beta}^{(k+1)}|k \sim TN(\boldsymbol{\mu}_0^{(k,k+1)}, \boldsymbol{\Sigma}_0^{(k,k+1)}|b^{(k+1)} > l_k)$ and $\sigma^{2(s)}|k \sim IG(\alpha_1^{(k,s)}, \alpha_2^{(k,s)})$ for $s = 1, \dots, k+1$, then $(\boldsymbol{\beta}_t|y_{1:t}, \sigma_t^2, \tau_t, s_t = s, k) \sim$

$$\begin{cases} N(\boldsymbol{\mu}_t^{(k,s)}, \boldsymbol{\Sigma}_t^{(k,s)}), & \text{if } s \leq k \\ TN(\boldsymbol{\mu}_t^{(k,k+1)}, \boldsymbol{\Sigma}_t^{(k,k+1)}|b_t^{(k+1)} > l_k), & \text{if } s = k+1 \end{cases} \quad (24)$$

$$(\sigma_t^2|y_{1:t}, \boldsymbol{\beta}_t, \tau_t, s_t = s, k) \sim IG\left(\alpha_1^{(k,s)} + \frac{t - \tau_t}{2}, \alpha_2^{(k,s)} + \frac{\|y_{\tau_t+1:t}^T - \mathbf{X}_{\tau_t+1:t}\boldsymbol{\beta}_t\|^2}{2}\right) \quad (25)$$

where

$$\boldsymbol{\mu}_t^{(k,s)} = \left[\frac{\mathbf{X}_{\tau_t+1:t}^T \mathbf{X}_{\tau_t+1:t}}{\sigma_t^2} + \boldsymbol{\Sigma}_0^{-1(k,s)} \right]^{-1} \left[\frac{\mathbf{X}_{\tau_t+1:t}^T y_{\tau_t+1:t}^T}{\sigma_t^2} + \boldsymbol{\Sigma}_0^{-1(k,s)} \boldsymbol{\mu}_0^{(k,s)} \right], s = 1, \dots, k+1 \quad (26)$$

and

$$\boldsymbol{\Sigma}_t^{(k,s)} = \left[\frac{\mathbf{X}_{\tau_t+1:t}^T \mathbf{X}_{\tau_t+1:t}}{\sigma_t^2} + \boldsymbol{\Sigma}_0^{-1(k,s)} \right]^{-1}, s = 1, \dots, k+1 \quad (27)$$

The Gibbs “move” step could effectively diversify particles and generate more particles with significant weights. However, the introduction could result in extra computational cost as well as break the balance of the computational load at each time step. Based on Lemma 1, all the observations from the latest change-point to the current time are used for Gibbs move. It is intuitive that the longer the phase duration, the higher the computational cost the Gibbs move will take. To control the computational cost, we adopt the “partial move” strategy [37], where randomly drawn particles among the resampled particles are moved until the sum of their durations $t - \tau_t^{(i)}$ is larger than a controlling constant C .

Although the Gibbs move step has solved the particle impoverishment issue for the continuous components $(\boldsymbol{\beta}_t, \sigma_t^2)$, it could not handle the same problem with the discrete component k of the state vector $\mathbf{x}_t = (\boldsymbol{\theta}^{(st)}, \tau_t, s_t, k)$. Indeed, the discrete component k of each particle is generated at the first time step and kept constant across all the following time steps. That means the impoverishment issue is much worse than the other components of the state vector. As we observed, after only several iterations, there may be only one unique k among all particles, which will in a totally failed PF algorithm. To solve this problem, we propose to use a stratified approach. Specifically, for each k_i or category i in the categorical distribution, the developed particle filtering algorithm with the same number of particles N is applied individually. In the posterior approximation, the extra group weight coefficient $W_t^{(g)}$ is applied to each category g or all particles of each group. The group weight coefficient $W_t^{(g)}$ can be calculated as

$$W_t^{(g)} = \frac{\sum_{i=1}^N \omega_t(\mathbf{x}_{1:t}^{(g,i)})}{\sum_{g=1}^{\dim(k)} \sum_{i=1}^N \omega_t(\mathbf{x}_{1:t}^{(g,i)})} \quad (28)$$

This strategy can effectively avoid the disappearing of certain k in the resampling process. We call this approach the *stratified particle filtering (SPF)*.

In summary, the developed SPF algorithm for sequential model updating is given in Algorithm 2.

Algorithm 2. Stratified Particle Filtering Algorithm for Sequential Model Updating

At time $t = 1$:

For $g = 1: \dim\{k\}$

1. Set $k^{(g,i)} = k_g$, Sample $\beta_1^{(g,i)} \sim N(\mu_0^{(k^{(g,i),1})}, \Sigma_0^{(k^{(g,i),1})})$, $\sigma_1^{2(g,i)} \sim IG(\alpha_1^{(k^{(g,i),1})}, \alpha_2^{(k^{(g,i),1})})$. Set $\tau_1^{(g,i)} = 0, s_1^{(g,i)} = 1$, and set $\mathbf{x}_1^{(g,i)} = (\beta_1^{(g,i)}, \sigma_1^{2(g,i)}, \tau_1^{(g,i)}, s_1^{(g,i)}, k^{(g,i)})$ for $i = 1:N$
2. Compute weights $\omega_1(\mathbf{x}_1^{(g,i)})$ and normalized weights $W_1^{(g,i)} = \omega_1(\mathbf{x}_1^{(g,i)}) / \sum_{i=1}^N \omega_1(\mathbf{x}_1^{(g,i)})$ based on Eq. (23) for $i = 1:N$
3. Resample $\{W_1^{(g,i)}, \mathbf{x}_1^{(g,i)} | i = 1, \dots, N\}$ according to their weights $W_1^{(g,i)}$ to obtain N equally weighted particles $\{\frac{1}{N}, \bar{\mathbf{x}}_1^{(g,i)}\}$ and set $\{W_1^{(g,i)}, \mathbf{x}_1^{(g,i)}\} \leftarrow \{\frac{1}{N}, \bar{\mathbf{x}}_1^{(g,i)}\}$

End

4. Calculate the group weight $W_1^{(g)}$ based on Eq. (28)
5. Set $W_1^{(g,i)} = \frac{W_1^{(g)}}{N}$ for $g = 1: \dim\{k\}$ and $i = 1:N$

At time $t \geq 2$:

For $g = 1: \dim\{k\}$

1. Calculate the probability $p_{t|t-1}^{(g,i)}$ based on Eq. (12)
2. Sample $u^{(g,i)} \sim U(0,1)$
 - If $u^{(g,i)} \leq p_{t|t-1}^{(g,i)}$,
 - i. Set $\tau_t^{(g,i)} = t - 1, s_t^{(g,i)} = s_{t-1}^{(g,i)} + 1$,
 - ii. Sample $\beta_t^{(g,i)} \sim N(\mu_0^{(k^{(g,i),s_t^{(g,i)}})}, \Sigma_0^{(k^{(g,i),s_t^{(g,i)}})})$ if $s_t^{(g,i)} \leq k^{(g,i)}$, otherwise, sample $\beta_t^{(g,i)} \sim TN(\mu_0^{(k^{(g,i),s_t^{(g,i)}})}, \Sigma_0^{(k^{(g,i),s_t^{(g,i)}})}) | b_t^{(g,i)} > l_g$
 - iii. Sample $\sigma_t^{2(g,i)} \sim IG(\alpha_1^{(k^{(g,i),s_t^{(g,i)}})}, \alpha_2^{(k^{(g,i),s_t^{(g,i)}})})$
 - iv. Set $\mathbf{x}_t^{(g,i)} \leftarrow (\beta_t^{(g,i)}, \sigma_t^{2(g,i)}, \tau_t^{(g,i)}, s_t^{(g,i)}, k^{(g,i)})$
 - Otherwise, set $\mathbf{x}_t^{(g,i)} \leftarrow \mathbf{x}_{t-1}^{(g,i)}$
3. Compute $\omega_t(\mathbf{x}_t^{(g,i)})$ and normalized weights $W_t^{(g,i)}$ based on Eq. (23).
4. Resample $\{W_t^{(g,i)}, \mathbf{x}_{1:t}^{(g,i)}\}$ to obtain N equally weighted particles $\{\frac{1}{N}, \bar{\mathbf{x}}_t^{(g,i)}\}$ and set $\{W_t^{(g,i)}, \mathbf{x}_t^{(g,i)}\} \leftarrow \{\frac{1}{N}, \bar{\mathbf{x}}_t^{(g,i)}\}$
5. Gibbs move: select a subset S of $\{\mathbf{x}^{(g,i)} | i = 1:N\}$ such that $\sum_{j \in S} (t - \tau^{(g,j)}) \leq C$
 - Sample $\beta_t^{(g,j)}$ based on Eq. (24)
 - Sample $\sigma_t^{2(g,j)}$ based on Eq. (25)

End

6. Calculate the group weight $W_t^{(g)}$ based on Eq. (28)
 7. Set $W_t^{(g,i)} = \frac{W_t^{(g)}}{N}$ for $g = 1: \dim\{k\}$ and $i = 1:N$
-

C. RUL Prediction

After the degradation model of the working unit is updated using the observations up to the current time, the next step is to predict the future degradation magnitude and RUL for preventive maintenance. Due to the multiple change-point that may occur in future, the exact Bayesian inference is intractable,

even if the current model is known. However, through the PF algorithm, the RUL prediction is proven to be very convenient.

Denote R_t as the RUL at the current time t . Then the distribution of R_t can be expressed by

$$\begin{aligned} P(R > L | y_{1:t}) &= P(y_{t+1} < \Gamma, y_{t+2} < \Gamma, \dots, y_{t+L} < \Gamma | y_{1:t}) \\ &= \int P(y_{t+1} < \Gamma, \dots, y_{t+L} < \Gamma | \mathbf{x}_{t+1:t+L}) f(\mathbf{x}_{t+L} | y_{1:t}) d\mathbf{x}_{t+L} \\ &= \int \left[\prod_{j=1}^L P(y_{t+j} < \Gamma | \mathbf{x}_{t+j}) f(\mathbf{x}_{t+j} | \mathbf{x}_{t+j-1}) \right] f(\mathbf{x}_t | y_{1:t}) d\mathbf{x}_{t:t+L} \end{aligned} \quad (29)$$

where Γ is the failure threshold. The above equation is not tractable analytically. However, we can conveniently generate samples from the distribution $f(\mathbf{x}_{t:t+L} | y_{1:t})$ based on the particles at the current time and the prior state transition process. Given the particles and their weights at the current time $\{W_t^{(g,i)}, \mathbf{x}_t^{(g,i)}, g = 1: \dim(k), i = 1:N\}$ which approximately follow $f(\mathbf{x}_t | y_{1:t})$, the samples of the future states $\{\mathbf{x}_{t+1:t+L}^{(i)}, i = 1, \dots, N\}$ can be generated through the prior state transition function $f(\mathbf{x}_{t+j} | \mathbf{x}_{t+j-1})$ given in Equation (10). Based on the generated samples, the RUL distribution can be approximated by

$$\begin{aligned} P(R > L | y_{1:t}) &= P(y_{t+1} < \Gamma, \dots, y_{t+L} < \Gamma | y_{1:t}) \\ &\approx \sum_{g=1}^{\dim(k)} \frac{W_t^{(g)}}{N} \sum_{i=1}^N \prod_{j=1}^L P(y_{t+j} < \Gamma | \mathbf{x}_{t+j}^{(g,i)}) \end{aligned} \quad (30)$$

Conditioning that there are k line segments or phases for a working unit, the failure will not occur before the k -th phase. Therefore, the probability $P(y_{t+j} < \Gamma | \mathbf{x}_{t+j}^{(g,i)})$ can be calculated by

$$\begin{aligned} P(y_{t+j} < \Gamma | \mathbf{x}_{t+j}^{(g,i)}) &= \begin{cases} \Phi(\Gamma | a_{t+j}^{(g,i)} + b_{t+j}^{(g,i)}(t+j), \sigma_t^{2(g,i)}), & s_t^{(g,i)} = k^{(g,i)} + 1 \\ 1, & s_t^{(g,i)} \leq k^{(g,i)} \end{cases} \end{aligned} \quad (31)$$

where $\Phi(\cdot)$ is the CDF of Gaussian distribution.

V. CASE STUDIES

In this section, the proposed method is evaluated through numerical simulations and real vibration data of rotational bearings. For all of case studies, we choose 50%, 70% and 90% of actual failure time as our starting points of RUL prediction.

A. Simulation Study

In this subsection, we evaluate the performance of the proposed method through simulated piecewise linear signals. For simplicity we assume that there are only two categories of degradation signals: two-phase and three-phase cases. The categorical distribution is given by

TABLE I
HYPER-PARAMETER SPECIFICATION FOR NUMERICAL SIMULATION

Variables	Two-phase Model	Three-phase Model
$\delta^{(s)}$	$\delta_0^{(1,1)} = 400, \sigma_0^{2(1,1)} = 225$ $\delta_0^{(1,2)} = 500, \sigma_0^{2(1,2)} = 100$	$\delta_0^{(2,1)} = 200, \sigma_0^{2(2,1)} = 100$ $\delta_0^{(2,2)} = 400, \sigma_0^{2(2,2)} = 100$ $\delta_0^{(2,3)} = 500, \sigma_0^{2(2,3)} = 100$
$\beta^{(s)}$	$\mu_0^{(1,1)} = [-15; 0.008]$ $\Sigma_0^{(1,1)} = \begin{bmatrix} 0.015 & 0.0014 \\ 0.0014 & 0.00046 \end{bmatrix}$ $\mu_0^{(1,2)} = [-30; 0.3]$ $\Sigma_0^{(1,2)} = \begin{bmatrix} 0.024 & -0.0007 \\ -0.0007 & 0.0057 \end{bmatrix}$	$\mu_0^{(2,1)} = [-10; 0.0005]$ $\Sigma_0^{(2,1)} = \begin{bmatrix} 0.15 & 0.00014 \\ 0.00014 & 0.0009 \end{bmatrix}$ $\mu_0^{(2,2)} = [-18; 0.02]$ $\Sigma_0^{(2,2)} = \begin{bmatrix} 0.024 & -0.0007 \\ -0.0007 & 0.000048 \end{bmatrix}$ $\mu_0^{(2,3)} = [-50; 0.08]$ $\Sigma_0^{(2,3)} = \begin{bmatrix} 0.075 & -0.00008 \\ -0.00008 & 0.00025 \end{bmatrix}$
$\sigma^{2(s)}$	$\alpha_1^{(1,1)} = 1.4, \alpha_2^{(1,1)} = 2.5$ $\alpha_1^{(1,2)} = 2, \alpha_2^{(1,2)} = 4$	$\alpha_1^{(2,1)} = 3.64, \alpha_2^{(2,1)} = 2$ $\alpha_1^{(2,2)} = 0.6, \alpha_2^{(2,2)} = 0.5$ $\alpha_1^{(2,3)} = 3.6, \alpha_2^{(2,3)} = 5$

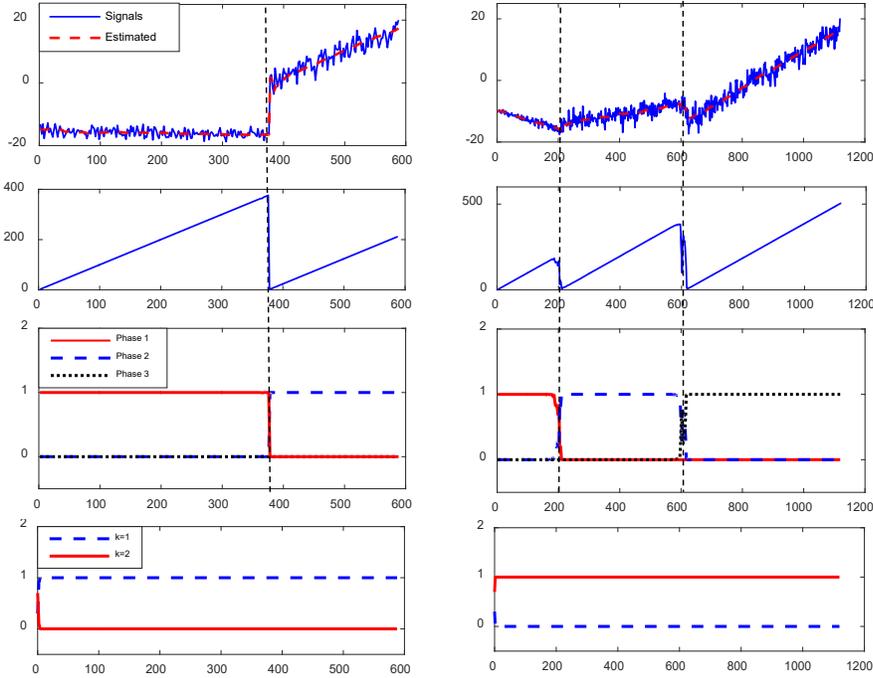


Figure 3. Illustration of the SPF based online monitoring of degradation signals with two phases (left panel) and three phases (right panel). (a) and (e): degradation signals and estimated signals; (b) and (f): the estimated duration of the current linear phase; (c) and (g): the probability mass function of the current phase; (d) and (h): the probability mass function of category. The vertical dashed lines are the true change-points.

$$k = \begin{cases} 1, & \text{with probability } p = 0.3 \\ 2, & \text{with probability } p = 0.7 \end{cases}$$

We assume the unit will fail once the observation reaches the threshold $\Gamma = 20$. The slope lower bound of last phase is set to be $l_1 = l_2 = 0.003$, The hyper-parameters of $\delta^{(k,s)}$, $\beta^{(k,s)}$ and $\sigma^{2(k,s)}$ are specified in TABLE I.

In total 200 CM signals are simulated, among which 69 are two-phase signals and 131 are three-phase signals. The BIC based model selection method can accurately obtain the right number of change-points and their locations for each simulated signal. Due to page limitation, the estimated hyper-parameters are not listed here. In the stratified particle filtering algorithm,

the number of particles for each category is set to be $N = 5000$. Figure 3 shows the online monitoring of degradation signals with one and two change-points. From the top two panels we can see that the estimated signals (mean value) are very close to the true values. The second row of these panels shows the mean value of the current phase length. As we can see, the algorithm can rapidly detect the phase change. The bottom four panels show the probability mass function of the discrete components (s_t, k) of the state vector. As we can see, the algorithm can accurately detect the number of phases the degradation signal will have and the current phase the degradation signal is at.

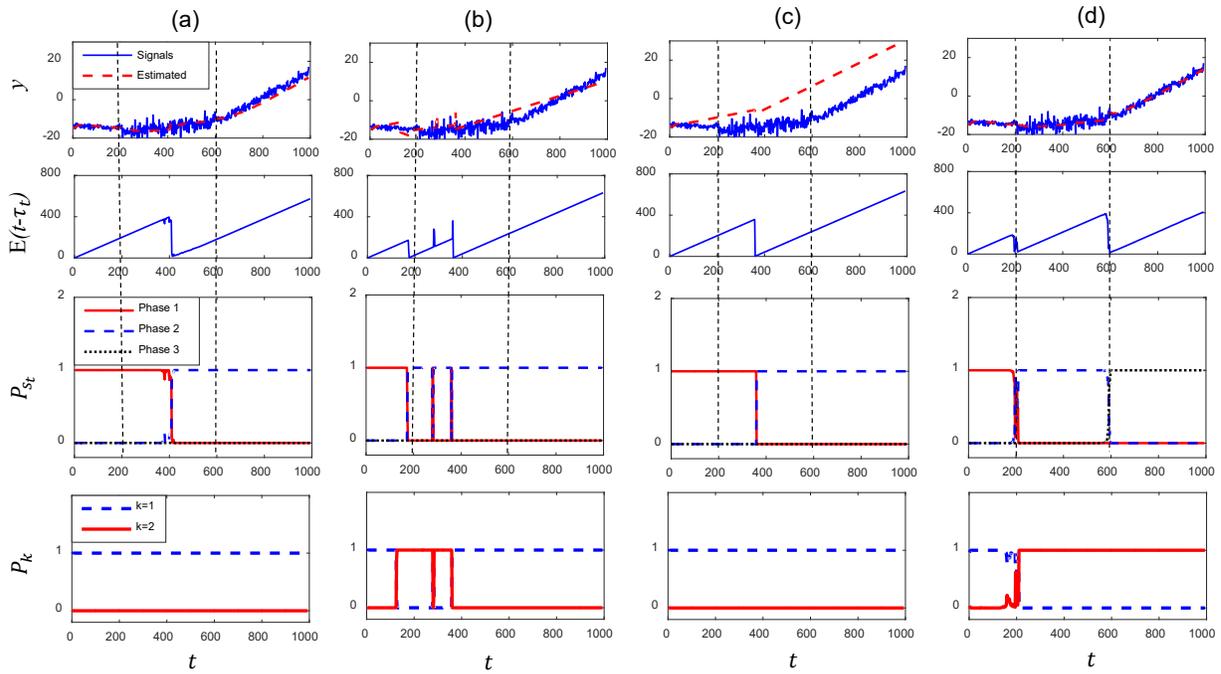


Figure 4. Comparison of the proposed SPF algorithm to other PF algorithms without either partial Gibbs move or stratified strategy: (a) no stratification, Gibbs move; (b) stratification, no Gibbs move; (c) no stratification, no Gibbs move and (d) proposed SPF with both particle Gibbs move and stratified approach.

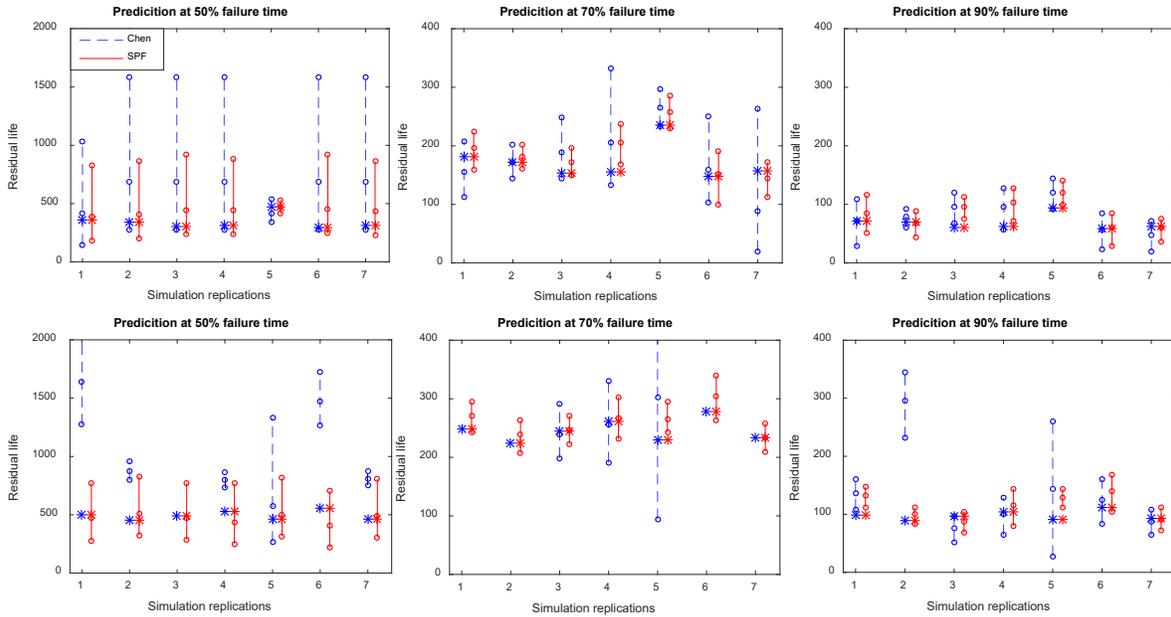


Figure 5. Prediction intervals of 7 simulated CM signals. The \circ represents the 5%,50%,95% quantiles of the RUL distributions, * is the actual RUL.(a)-(c) Prediction intervals for two-phase signals; (d)-(f) Prediction intervals for three-phase signals.

Figure 4 shows the comparison of the SPF algorithm to three other PF algorithms without either partial Gibbs move or stratified strategy. The number of particles here is set to be 500. Clearly, without the stratified strategy (a and c), all the particles with discrete component $k = 2$ gradually diminish along iterations, which results in an inaccurate model with only two phases ($k = 1$). Without the partial Gibbs move (b and c), the degeneracy of the continuous components occurs, which significantly influence the model accuracy. The proposed SPF algorithm has effectively overcome the particle degeneracy and

impoverishment issues and works quite well with only 500 samples.

To evaluate the prediction performance, 200 new degradation signals are randomly generated as a testing dataset using the specified priors. We compare the SPF algorithm with Chen’s two-phase model [15], where only one change-point is considered. For Chen’s method, all the 200 training dataset with both two-phase and three-phase signals are used to estimate the priors of the two-phase model. Figure 5 shows the comparison of the prediction intervals of 7 simulated two-phase signals and 7 three-phase signals predicted at 50%,

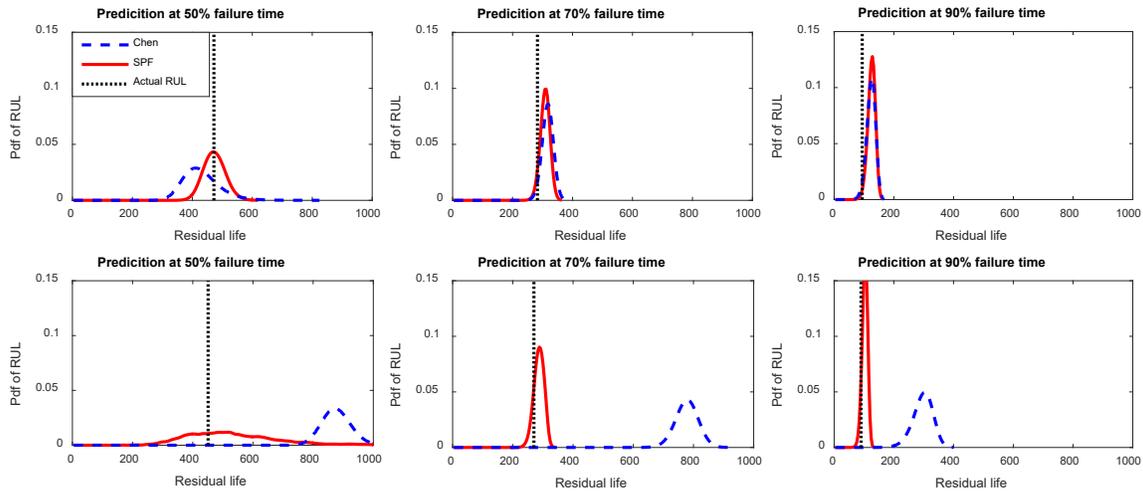


Figure 6. Comparison of the pdf of the RUL. (a)-(c) two-phase signal (d)-(f) three-phase signal.

70% and 90% of actual failure time. Figure 6 shows the detailed RUL prediction of the 5th and the 2nd unit of the seven signals of each category in Figure 5. From Figure 5 we can see that our prediction accuracy is much better than Chen’s method in almost all the 14 cases. For two-phase signals, both methods work well. However, our method is slightly better at 70% and 90% of the failure time while much better at 50% of the failure time than Chen’s method, which can also be seen from Figure 6 (a-c). The main reason is that in Chen’s method, the priors are estimated using all two-phase and three-phase signals, which will result in less accurate priors. At the 50% failure time, the prediction accuracy is mainly determined by the prior knowledge, while at the 70% and 90% of the failure time, the observations dominate the posterior distributions. Therefore at the early stage, our method with more accurate priors is much better than Chen’s method while at the late stage, the performances of both methods are comparable. For three-phase signals, our method is much better than Chen’s method at all the three prediction times, as shown in Figure 5 (d-f) and Figure 6 (d-f). It is expected since the two-phase model is inadequate to model signals with three phases.

To evaluate the overall performance, we use the root-mean-square-deviation (RMSD), which is defined as $RMSD = \sqrt{E(R - R_{true})^2}$, where R and R_{true} are the predicted and true RUL respectively. Since the proposed method is a Monte Carlo based method, there exists inevitable randomness (though very small). So for each signal the SPF algorithm is repeated 10 times. TABLE II shows the RMSD of the proposed method and Chen’s method using the 200 testing signals. As we can see, the proposed method is much more accurate than Chen’s method, with the RMSD reduced by more than 70% at almost all six prediction times. As the prediction time approaches to the true failure time, the RMSD of the proposed method monotonically decreases. This is highly desirable since it becomes more and more important to get an accurate prediction when the RUL approaches zero. However, for Chen’s method, RMSD first increases and then decreases. That means the prediction error at the second phase is even worse than making prediction at the first phase. The reason is that for

three-phase signals, the second phase with a relative small degradation rate is detected as the final phase in Chen’s method. The more observations in the second-phase, the flatter the final phase of the updated model and thus the worse the prediction.

TABLE II
COMPARISON OF THE RMSD AT SIX PREDICTION TIMES

Method	RMSD					
	40%	50%	60%	70%	80%	90%
Chen	593.4	649.4	696.7	548.7	188.6	58.6
SPF	222.1	194.6	133.8	74.4	23.4	21.9

The computational costs of the SPF and Chen’s method using MATLAB running on an i7-6560U 2.21 GHz Intel processor are shown in TABLE III. For the SPF method, the total number of particles is set 5000. In the model updating stage, we compare the computational costs of these two methods running 300 time steps. For the prediction stage, the costs of running different time steps are evaluated, since the cost of prediction in Chen’s method nonlinearly increases with time steps. As we can see, the SPF method is much more expensive in the model updating stage than Chen’s method. However, in the prediction stage, the cost of Chen’s method exponentially increases with the time steps, due to the CDF computation of a multivariate t distribution with an increasing dimension. For the SPF method, the computational cost of the prediction linearly increases with the time steps. Note that the selection of 5000 particles is quite conservative. From Figure 4 we can see that the model updating is quite accurate with only 500 particles.

TABLE III
COMPARISON OF THE COMPUTATIONAL COST (UNIT: SECONDS)

Method	Updating		Prediction			
	300	25	50	75	100	125
Chen	0.024	2.1	7.4	16.1	28.8	45.7
SPF	179.4	6.2	11.7	16.9	22.0	27.0

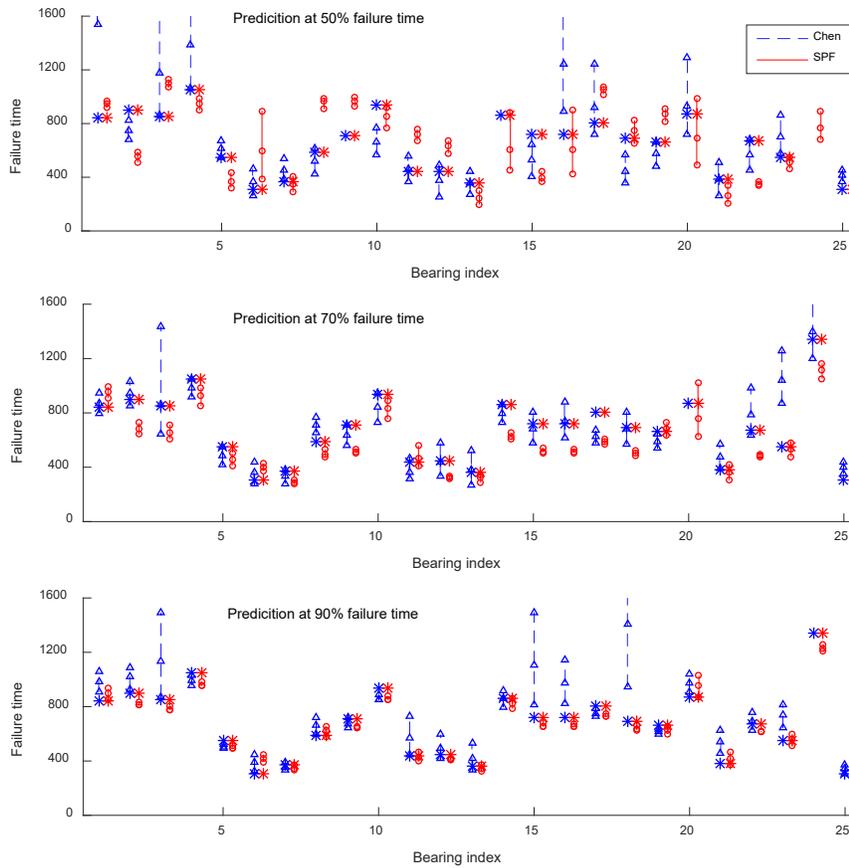


Figure 7. Prediction intervals of the 25 bearing signals

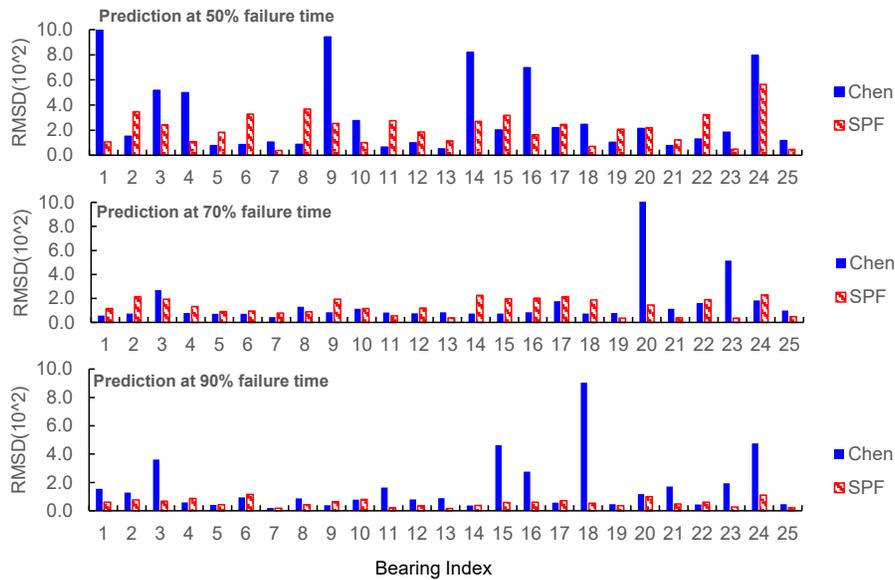


Figure 8. RMSD of the 25 bearing signals

B. Degradation Monitoring of the Rotational Bearings

In this section, the proposed method is applied to the real degradation signals of rotational bearings [6, 10, 15]. They are vibrational signals (log-transformed) of a set of identical thrust ball bearings captured by an accelerometer in the accelerated aging testing. There are in total 25 complete bearing signals

available. The data sampling interval is 2 minutes for each signal. When the vibration magnitude exceeds the threshold $\Gamma = \log(0.03)$, which was computed from published industrial standards, the bearing is considered failed [6].

In the offline modeling and prior estimation process, we set the maximum number of phases for all 25 signals to be 3 to control the model complexity. It is found that all bearing signals

TABLE IV
ESTIMATED HYPER-PARAMETERS OF THE PRIOR DISTRIBUTIONS

	$s = 1$	$s = 2$	$s = 3$
$\delta^{(s)}$	$\delta_0^{(1)} = 246, \sigma_0^{2(1)} = 163^2$	$\delta_0^{(2)} = 199, \sigma_0^{2(2)} = 123^2$	$\delta_0^{(3)} = 232, \sigma_0^{2(3)} = 123^2$
$\beta^{(s)}$	$\mu_0^{(1)} = [-7.14, 0.00027]$ $\Sigma_0^{(1)} = \begin{bmatrix} 0.15 & -0.0003 \\ -0.0003 & 1.30 \times 10^{-5} \end{bmatrix}$	$\mu_0^{(2)} = [-6.42, 0.0028]$ $\Sigma_0^{(2)} = \begin{bmatrix} 2.05 & -0.009 \\ -0.009 & 7.36 \times 10^{-5} \end{bmatrix}$	$\mu_0^{(3)} = [-6.8, 0.005]$ $\Sigma_0^{(3)} = \begin{bmatrix} 3.89 & -0.008 \\ -0.008 & 2.52 \times 10^{-5} \end{bmatrix}$ $l = 0.005$
$\sigma^{2(s)}$	$\alpha_1^{(1)} = 2.65, \alpha_2^{(1)} = 0.01$	$\alpha_1^{(2)} = 0.54, \alpha_2^{(2)} = 0.004$	$\alpha_1^{(3)} = 1.28, \alpha_2^{(3)} = 0.03$

with three phases have the minimum BIC. The estimated hyper-parameters of the prior distributions are shown in TABLE IV. It is observable that the slope is quite small at the first phase, indicating a stable operation process. The slopes of the following phases are larger than the former ones. It indicates that when a new change-point occurs, the degradation rate of bearings increases.

Figure 7 shows the prediction intervals at 50%, 70% and 90% of failure time against the actual failure time. We can see that the prediction intervals at 90% failure time are much narrower than that of 70% failure time and 50% failure time. As the prediction time is closer to actual failure time, the intervals become smaller. It is obvious that the more observed data, the more accurate the prediction. Figure 8 shows the RMSD of the 25 bearing signals. Compared with Chen's method, the predictive accuracy of the proposed method is significantly improved. TABLE V shows the comparison of the SPF method with Chen's method and the GLLR method [6] in terms of RMSD at the three time steps. In the GLLR method, the first phase with normal working condition is manually truncated first, and the remaining data are fitted using Bayesian simple linear regression. Clearly, our method outperforms these two methods at all three prediction times. The GLLR method has the largest prediction error on this dataset.

TABLE V
COMPARISON OF THE SPF METHOD WITH GLLR AND CHEN'S METHOD

Method	RMSD		
	50%	70%	90%
GLLR	356.5	234.2	227.8
Chen	318.4	156.9	169.4
SPF	210.0	131.7	56.9

VI. CONCLUSION AND DISCUSSION

In this paper we propose a multiple-phase modeling of degradation signals for health condition monitoring and remaining useful life prediction. To integrate the historical data with in-situ observations of each new unit in the RUL prediction, the multiple change-point model is formulated under the Bayesian framework and a novel stochastic process is proposed as priors of the formulated model. To facilitate the online monitoring and RUL prediction, the multiple change-point model is first represented by a novel nonstandard state-space model and then a new particle filtering algorithm is developed for online model updating and RUL prediction. A stratified sampling approach and a partial Gibbs resample-move strategy are developed to overcome the particle

impoverishment problem and reduce the computational burden. The advantages of the proposed method have been demonstrated through extensive numerical studies and real case studies.

Nevertheless, there still exist several open issues that need to be investigated. First, in the proposed method, all phases are assumed independent in the prior specification. However, in practice all phases are often connected and highly correlated. Incorporating the phase correlation may improve the prior informativeness and thus improve the prediction accuracy. Second, the computational cost of SPF algorithm may be higher than most of other existing methods, which may prohibit its applications where rapid prediction is required. These issues will be left to our future work.

APPENDIX

Proof of Lemma 1

For the sake of simplicity, we ignore all the superscripts (k and s) in the proof.

(1) If $\beta_t \sim N(\mu_0, \Sigma_0)$, then

$$\begin{aligned} p(\beta_t | y_{1:t}, \sigma_t^2, \tau_t) &\propto p(\beta_t) p(y_{\tau_t+1:t} | \tau_t, \sigma_t^2, \beta_t) \\ &\propto \exp \left[-\frac{(\beta_t - \mu_0)^T \Sigma_0^{-1} (\beta_t - \mu_0)}{2} \right] \cdot \exp \left(-\frac{\|y_{\tau_t+1:t}^T - X_{\tau_t+1:t} \beta_t\|^2}{2\sigma_t^2} \right) \\ &\propto \exp \left\{ -\frac{1}{2} \left[\beta_t^T \left(\frac{X_{\tau_t+1:t}^T X_{\tau_t+1:t}}{\sigma_t^2} + \Sigma_0^{-1} \right) \beta_t \right. \right. \\ &\quad \left. \left. - 2 \left(\mu_0^T \Sigma_0^{-1} + \frac{y_{\tau_t+1:t} X_{\tau_t+1:t}}{\sigma_t^2} \right) \beta_t \right] \right\} \\ &\propto \exp \left[-\frac{1}{2} (\beta_t - \mu_t)^T \Sigma_t^{-1} (\beta_t - \mu_t) \right] \end{aligned}$$

where

$$\begin{aligned} \mu_t &= \left[\frac{X_{\tau_t+1:t}^T X_{\tau_t+1:t}}{\sigma_t^2} + \Sigma_0^{-1} \right]^{-1} \left[\frac{X_{\tau_t+1:t}^T y_{\tau_t+1:t}}{\sigma_t^2} + \Sigma_0^{-1} \mu_0 \right] \\ \Sigma_t &= \left[\frac{X_{\tau_t+1:t}^T X_{\tau_t+1:t}}{\sigma_t^2} + \Sigma_0^{-1} \right]^{-1} \end{aligned}$$

Therefore

$$(\beta_t | y_{1:t}, \sigma_t^2, \tau_t) \sim N(\mu_t, \Sigma_t)$$

If $\beta_t \sim TN(\mu_0, \Sigma_0 | b_t > l)$, then similarly,

$$\begin{aligned} p(\beta_t | y_{1:t}, \sigma_t^2, \tau_t) &\propto p(\beta_t) p(y_{\tau_t+1:t} | \tau_t, \sigma_t^2, \beta_t) \\ &\propto \exp \left[-\frac{(\beta_t - \mu_0)^T \Sigma_0^{-1} (\beta_t - \mu_0)}{2} \right] \cdot I_{(l, +\infty)}(b_t) \\ &\quad \cdot \exp \left(-\frac{\|y_{\tau_t+1:t}^T - X_{\tau_t+1:t} \beta_t\|^2}{2\sigma_t^2} \right) \end{aligned}$$

$$\propto \exp \left[-\frac{(\boldsymbol{\beta}_t - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\beta}_t - \boldsymbol{\mu}_0)}{2} \right] \cdot I_{(l,+\infty)}(b_t) \cdot \exp \left(-\frac{\|y_{\tau_t+1:t}^T - \mathbf{X}_{\tau_t+1:t} \boldsymbol{\beta}_t\|^2}{2\sigma_t^2} \right)$$

$$\propto \exp \left[-\frac{1}{2} (\boldsymbol{\beta}_t - \boldsymbol{\mu}_t)^T \boldsymbol{\Sigma}_t^{-1} (\boldsymbol{\beta}_t - \boldsymbol{\mu}_t) \right] \cdot I_{(l,+\infty)}(b_t)$$

where $I_{(l,+\infty)}(\cdot)$ is an indicator function, i.e., $I_{(l,+\infty)}(x) = 0$ if $x \leq l$ and $I_{(l,+\infty)}(x) = 1$ if $x > l$

Therefore

$$(\boldsymbol{\beta}_t | y_{1:t}, \sigma_t^2, \tau_t) \sim TN(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t | b_t > l)$$

$$(2) p(\sigma_t^2 | \boldsymbol{\beta}_t, y_{1:t}, \tau_t) \propto p(\sigma_t^2) p(y_{\tau_t+1:t} | \boldsymbol{\beta}_t, \tau_t, \sigma_t^2)$$

$$\propto (\sigma_t^2)^{-(\alpha_1+1)} \exp \left(-\frac{\alpha_2}{\sigma_t^2} \right) (\sigma_t^2)^{-\frac{t-\tau_t}{2}}$$

$$\exp \left(-\frac{\|y_{\tau_t+1:t}^T - \mathbf{X}_{\tau_t+1:t} \boldsymbol{\beta}_t\|^2}{2\sigma_t^2} \right)$$

$$\propto (\sigma_t^2)^{-(\alpha_1 + \frac{t-\tau_t}{2} - 1)} \exp \left(-\frac{\alpha_2 + \|y_{\tau_t+1:t}^T - \mathbf{X}_{\tau_t+1:t} \boldsymbol{\beta}_t\|^2 / 2}{\sigma_t^2} \right)$$

Therefore

$$(\sigma_t^2 | \boldsymbol{\beta}_t, y_{1:t}, \tau_t) \sim IG \left(\alpha_1 + \frac{t-\tau_t}{2}, \alpha_2 + \frac{\|y_{\tau_t+1:t}^T - \mathbf{X}_{\tau_t+1:t} \boldsymbol{\beta}_t\|^2}{2} \right)$$

REFERENCES

- [1] 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.
- [2] R. Kothamasu, S. H. Huang, and W. H. VerDuin, "System health monitoring and prognostics—a review of current paradigms and practices," in *Handbook of Maintenance Management and Engineering*, ed: Springer, 2009, pp. 337-362.
- [3] A. H. Elwany and N. Z. Gebraeel, "Sensor-driven prognostic models for equipment replacement and spare parts inventory," *IIE Transactions*, vol. 40, pp. 629-639, 2008.
- [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] C. Okoh, R. Roy, J. Mehnen, and L. Redding, "Overview of remaining useful life prediction techniques in through-life engineering services," *Procedia CIRP*, vol. 16, pp. 158-163, 2014.
- [6] 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.
- [7] E. Zio and F. Di Maio, "A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system," *Reliability Engineering & System Safety*, vol. 95, pp. 49-57, 2010.
- [8] O. E. Dragomir, R. GOURTVEAU, N. Zerhouni, and F. Dragomir, "Framework for a distributed and hybrid prognostic system," *IFAC Proceedings Volumes*, vol. 40, pp. 431-436, 2007.
- [9] 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.
- [10] 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.
- [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] S. J. Bae and P. H. Kvam, "A nonlinear random-coefficients model for degradation testing," *Technometrics*, vol. 46, pp. 460-469, 2004.
- [13] C. J. Lu and W. O. Meeker, "Using degradation measures to estimate a time-to-failure distribution," *Technometrics*, vol. 35, pp. 161-174, 1993.
- [14] R. B. Chinnam, "On-line reliability estimation for individual components using statistical degradation signal models," *Quality and Reliability Engineering International*, vol. 18, pp. 53-73, 2002.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] Y. Li and P. Nikitsaranont, "Gas turbine performance prognostic for condition-based maintenance," *Applied energy*, vol. 86, pp. 2152-2161, 2009.
- [20] X.-S. Si, C.-H. Hu, X. Kong, and D.-H. Zhou, "A residual storage life prediction approach for systems with operation state switches," *IEEE Transactions on Industrial Electronics*, vol. 61, pp. 6304-6315, 2014.
- [21] X.-S. Si, C.-H. Hu, Q. Zhang, and T. Li, "An Integrated Reliability Estimation Approach With Stochastic Filtering and Degradation Modeling for Phased-Mission Systems," *IEEE transactions on cybernetics*, vol. 47, pp. 67-80, 2017.
- [22] Y. Wang, Y. Peng, Y. Zi, X. Jin, and K.-L. Tsui, "A Two-Stage Data-Driven-Based Prognostic Approach for Bearing Degradation Problem," *IEEE Transactions on Industrial Informatics*, vol. 12, pp. 924-932, 2016.
- [23] R. J. Meinhold and N. D. Singpurwalla, "Understanding the Kalman filter," *The American Statistician*, vol. 37, pp. 123-127, 1983.
- [24] A. Hannart and P. Naveau, "An improved Bayesian information criterion for multiple change-point models," *Technometrics*, vol. 54, pp. 256-268, 2012.
- [25] 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.
- [26] Y. Hou, J. Wu, and Y. Chen, "Online Steady State Detection Based on Rao-Blackwellized Sequential Monte Carlo," *Quality and Reliability Engineering International*, 2016.
- [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*, pp. 1-15, 2016.
- [28] G. Schwarz, "Estimating the dimension of a model," *The annals of statistics*, vol. 6, pp. 461-464, 1978.
- [29] R. Killick, P. Fearnhead, and I. Eckley, "Optimal detection of changepoints with a linear computational cost," *Journal of the American Statistical Association*, vol. 107, pp. 1590-1598, 2012.
- [30] B. E. Olivares, M. A. C. Munoz, M. E. Orchard, and J. F. Silva, "Particle-filtering-based prognosis framework for energy storage devices with a statistical characterization of state-of-health regeneration phenomena," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, pp. 364-376, 2013.
- [31] E. Zio and G. Peloni, "Particle filtering prognostic estimation of the remaining useful life of nonlinear components," *Reliability Engineering & System Safety*, vol. 96, pp. 403-409, 2011.
- [32] Q. Miao, L. Xie, H. Cui, W. Liang, and M. Pecht, "Remaining useful life prediction of lithium-ion battery with unscented particle filter technique," *Microelectronics Reliability*, vol. 53, pp. 805-810, 2013.
- [33] N. Li, Y. Lei, J. Lin, and S. X. Ding, "An improved exponential model for predicting remaining useful life of rolling element bearings," *IEEE Transactions on Industrial Electronics*, vol. 62, pp. 7762-7773, 2015.
- [34] Y. Hu, P. Baraldi, F. Di Maio, and E. Zio, "A particle filtering and kernel smoothing-based approach for new design component prognostics," *Reliability Engineering & System Safety*, vol. 134, pp. 19-31, 2015.
- [35] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and computing*, vol. 10, pp. 197-208, 2000.
- [36] W. R. Gilks and C. Berzuini, "Following a moving target—Monte Carlo inference for dynamic Bayesian models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, pp. 127-146, 2001.
- [37] N. Chopin, "Dynamic detection of change points in long time series," *Annals of the Institute of Statistical Mathematics*, vol. 59, pp. 349-366, 2007.



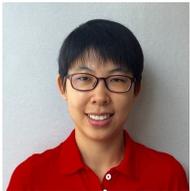
Yuxin Wen received the B.S. degree in Medical Informatics Engineering from Sichuan University, Sichuan, China in 2011, the M.S. degree in Biomedical Engineering from Zhejiang University, Zhejiang, China in 2014. Currently, she is pursuing the Ph.D. degree in Electrical and Computer Engineering at the University of Texas at El Paso, 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 at University of Texas-El Paso, 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, 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 the Institute for Operations Research and the Management Sciences (INFORMS), the Institute of Industrial and Systems Engineers (IISE), the Society of Manufacturing Engineers (SME).



Yuan Yuan is a Research Scientist at the IBM Research-- Singapore. She received her B.E. degree (2006) from Tsinghua University, Beijing, China, M.S. degrees in Industrial and Systems Engineering (2010) and Statistics (2011), and Ph.D. degree in Industrial and Systems Engineering (2014)

from University of Wisconsin-Madison, WI, USA. Her research mainly focuses on data analytics, in particular developing innovative and generic data-driven modelling and analysis methodologies for complex systems with massive data. She has received a number of awards including the QSR Best Student Paper Award from the Institute for Operations Research and the Management Sciences (INFORMS) (2014), the featured article award of IE magazine (2010).