

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

Yuxin Wen, Jianguo Wu, and 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 (CM), multiple change-point model, particle filters (PF), prognostics and health management, remaining useful life (RUL) prediction.

## NOMENCLATURE

### Abbreviations and Acronyms

RUL	Remaining useful life.
CM	Condition monitoring.
BIC	Bayesian information criterion.
i.i.d.	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 $i$ .

Manuscript received November 7, 2016; revised February 24, 2017 and April 26, 2017; accepted May 28, 2017. Date of publication June 15, 2017; date of current version August 30, 2017. Associate Editor: J. Liu. (Corresponding author: Jianguo Wu.)

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

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

Y. Yuan is with the IBM Research, Singapore 18983 (e-mail: polarisyy@gmail.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TR.2017.2710319

$c_{ik_i}$	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 $i$ .
$\sigma_{ik}$	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.
$\tau_t$	Latest change-point at time $t$ .
$\delta^{(s)}$	Duration of the $s$ th segment.
$\psi$	Hyperparameters of all priors.
$\theta^{(s)}$	Model parameters of the $s$ th segment.
$p_{t+1}(\mathbf{x}_t)$	Prior transition probability of state vector $\mathbf{x}_t$ at time $t + 1$ .
$\beta^{(s)}$	Vector of coefficients including slope and intercept.
$\hat{\beta}^{(s)}$	Estimated coefficients including slope and intercept.
$\hat{\sigma}^{2(s)}$	Estimated measurement noise.
$\mu_t^{(k,s)}$	Mean of normal distribution of the $s$ th segment.
$\Sigma_t^{(k,s)}$	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 the 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 $t$ .
$\Gamma$	Failure threshold.
$\Phi(\cdot)$	Cumulative distribution function of Gaussian distribution.

## I. INTRODUCTION

**T**HE 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 CM and RUL prediction in the past few decades [6]–[8]. According to Zio *et al.* [9] and Dragomir *et al.* [10], 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.* [11].

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 [9]. In these methods, the functional forms of the models are often linear, polynomial, exponential, or combination of them [12]. 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.* [13] 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 [14] 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 high-performance capacitors [15], the semiconductor laser diodes [16], and vibrational signals of rotational bearings [17]. 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 [18], [19]. 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.* [20] 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 [21] employed a combination of a linear model in the first phase and a quadratic model in the second phase to estimate the RUL 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 preset location of a change-point to illustrate the updating process of rolling element bearings. Later Chen and Tsui [17] 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 CM 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 [22], multiple consecutive phases of operations required to finish the service for phased-mission systems [23] and 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], [24]), or different patterns shown on CM signals, which may not have specific physical meanings [13], [17]. 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.* [22], [23] 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 offline 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 [17], [25]. However, there are several challenges on how to effectively apply the multiple change-point model for CM 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 [26], 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 offline modeling and online model updating and RUL prediction. In the offline modeling, a novel stochastic process is proposed to specify and estimate priors in the offline modeling. In the online stage, the multiple change-point model is formulated as a nonstandard state-space model and a novel stratified particle filtering (SPF) algorithm is developed for online model updating and RUL prediction. The contribution of this paper lies in the following threefold:

- 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 SPF 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 II, 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 III. Section IV 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 V demonstrates the effectiveness and accuracy of the proposed method through numerical and case studies. The conclusion and discussion are given in Section VI.

## 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 nonlinear 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.

Here we use the bearing vibration signals [6], [17] to demonstrate the superiority of the proposed method. As shown in Fig. 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 prespecified failure threshold. From Fig. 1(b) we can clearly see that a three-phase model is much more accurate. Note that

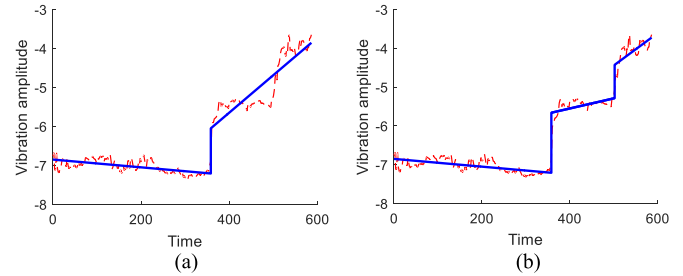


Fig. 1. Piecewise linear model for bearing signal. (a) One change-point. (b) Two change-points.

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  $\mathbf{Y}_i = \{y_{i1}, y_{i2}, \dots, y_{iT_i}\}$ . Suppose the  $k_i$  change-points are the integer-valued observation indices  $c_{i1}, c_{i2}, \dots, c_{ik_i}$ . For notational convenience, we define  $c_{i0} = 0$  and  $c_{ik_i+1} = T_i$ . Then  $c_{i0} = 0 < c_{i1} < c_{i2} < \dots < c_{ik_i} < c_{ik_i+1} = T_i$  and these  $k_i$  change-points divide the signal into  $k_i + 1$  line segments. Mathematically, the piece-wise linear model can be expressed as

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

where  $\varepsilon_{ij}$  follows i.i.d. standard normal distribution,  $a_{ik}$ ,  $b_{ik}$ , and  $\sigma_{ik}$  are the intercept, slope, and standard deviation of the  $k$ th line segment, respectively. Denote the true degradation signal without noise as  $\bar{y}_{ij}$ . 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  $a_t$ ,  $b_t$ ,  $\sigma_t^2$  are the parameters of the current line segment,  $\tau_t$  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 true degradation  $\bar{y}_{t^*}$ ,  $\forall t^* > t$  needs to be calculated based on the updated model

$$p(\bar{y}_{t^*} | y_{1:t}) = \sum_{k, s_t, \tau_t} \int p(\bar{y}_{t^*} | \mathbf{x}_t) p(\mathbf{x}_t | y_{1:t}) da_t db_t d\sigma_t^2 \quad (3)$$

where

$$p(\bar{y}_{t^*} | \mathbf{x}_t) = \sum_{s_{t^*}, \tau_{t^*}} \int p(\bar{y}_{t^*} | \mathbf{x}_{t^*}) \pi(\mathbf{x}_{t^*} | \mathbf{x}_t) da_{t^*} db_{t^*} d\sigma_{t^*}^2. \quad (4)$$

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 III 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 VI 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 Markov process could be assumed to simultaneously model the priors for the number of change-points and their occurrence intervals or equivalently their locations [27]. For example, a Poisson process could be used to model the occurrence of change-points, where the intervals  $\delta^{(s)}$  between successive change-points 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

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

where  $f(\cdot | \lambda)$  is the probability density function of an exponential distribution,  $\lambda$  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 intervals follow an i.i.d. geometric distribution [28]–[30]. The joint density is simply

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

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 noninformative, 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 modeled by a stochastic process where the first  $k$  interval lengths  $\{\delta^{(s)}\}_{s=1}^k$  follow independent and nonidentical distributions and the  $k+1$  model parameters  $\{\theta^{(s)}\}_{s=1}^{k+1}$  follow independent and nonidentical 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  $\pi(k = g) = p_g$  and  $\sum_g p_g = 1$ . For simplicity, we assume the phase durations follow normal distributions,  $\delta^{(s)} | k \sim N(\delta_0^{(k, s)}, \sigma_0^{2(k, s)})$ . For the changing parameters, the commonly used normal and inverse Gamma are specified, i.e.,  $\beta^{(s)} | k = (a^{(s)}, b^{(s)} | k)' \sim N(\boldsymbol{\mu}_0^{(k, s)}, \boldsymbol{\Sigma}_0^{(k, s)})$  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:  $\beta^{(k+1)} | k \sim TN(\boldsymbol{\mu}_0^{(k, k+1)}, \boldsymbol{\Sigma}_0^{(k, k+1)} | b^{(k+1)} > l_k)$  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 hyperparameters of all these priors, denoted as  $\psi$ , need to be estimated based on the historical data. One common way is to estimate  $\psi$  by maximizing the marginal likelihood [17] of  $I$  historical CM signals:

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

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 hyperparameters 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) [31] 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 ( $k$  change-points,  $k + 1$  slopes, intercepts and noise variance),  $l(\mathcal{M}|\mathbf{Y})$  is the log-likelihood given as

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

where

$$\mathbf{X}_{c_{s-1}+1:c_s} = \begin{bmatrix} \mathbf{1} & \mathbf{1} & \cdots & \mathbf{1} \\ \tau_{c_{s-1}+1} & \tau_{c_{s-1}+2} & \cdots & c_s \end{bmatrix}^T. \quad (8)$$

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

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

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 [32], 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 nonstandard state-space model with state vector  $\mathbf{x}_t = (\boldsymbol{\theta}^{(s_t)}, \tau_t, s_t, k)$  and prior state transition process

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

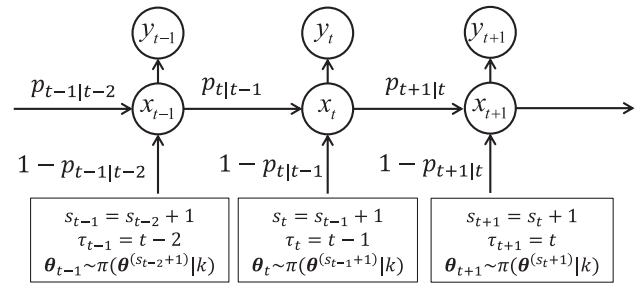


Fig. 2. Illustration of the formulated state-space model.

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

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

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.

The formulated state-space model is illustrated in Fig. 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 [33]; 2) linear state transition model, nonlinear observation model [34], [35]; and 3) nonlinear state transition model and nonlinear observation model [36], [37]. 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 SPF 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 methods to solve the filtering problems. It is particularly useful for sequential Bayesian inference of linear/nonlinear Gaussian/non-Gaussian state-space models [38]. 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

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

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

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

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, (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:

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

---

#### 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  $\{W_t^{(i)}, \mathbf{x}_{1:t}^{(i)}\} \leftarrow \{\frac{1}{N}, \bar{\mathbf{x}}_{1:t}^{(i)}\}$
- 

then the unnormalized weight function can be expressed by

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

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.

##### B. SPF 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 [38]. 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 (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 [28], 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 [39], 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  $(\beta_t, \sigma_t^2)$  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  $\beta^{(s)} | k = (a^{(s)}, b^{(s)} | k)' \sim N(\boldsymbol{\mu}_0^{(k,s)}, \boldsymbol{\Sigma}_0^{(k,s)})$  for  $s = 1, \dots, k$ ,  $\beta^{(k+1)} | k \sim TN(\boldsymbol{\mu}_0^{(k,k+1)}, \boldsymbol{\Sigma}_0^{(k,k+1)} | b^{(k+1)} > l_k)$  and  $\sigma^{2(s)} | k \sim IG(\alpha_1^{(k,s)}, \alpha_2^{(k,s)})$  for  $s = 1, \dots, k+1$ , then

$$\begin{aligned} & (\beta_t | y_{1:t}, \sigma_t^2, \tau_t, s_t = s, k) \sim \\ & \begin{cases} N(\boldsymbol{\mu}_t^{(k,s)}, \boldsymbol{\Sigma}_t^{(k,s)}), & \text{if } s \leq k \\ TN(\boldsymbol{\mu}_t^{(k,k+1)}, \boldsymbol{\Sigma}_t^{(k,k+1)} | b_t^{(k+1)} > l_k), & \text{if } s = k+1 \end{cases} \end{aligned} \quad (24)$$

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

where

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

and

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

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 [40], 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  $(\beta_t, \sigma_t^2)$ , it could not handle the same problem with the discrete component  $k$ . 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 value for  $k$  among all particles, which will result in a totally failed PF algorithm. To solve this problem, we propose to use a stratified approach. Specifically, for each category  $k = g$  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

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

This strategy can effectively avoid the disappearing of certain  $k$  in the resampling process. We call this approach the *SPF*.

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

### 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,

---

**Algorithm 2:** SPF 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
 
$$\tau_1^{(g,i)} = 0, s_1^{(g,i)} = 1, \text{ and set}$$

$$\mathbf{x}_1^{(g,i)} = (\beta_1^{(g,i)}, \sigma_1^{2(g,i)}, \tau_1^{(g,i)}, s_1^{(g,i)}, k^{(g,i)})$$
 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 (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 (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 (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
 
$$\beta_t^{(g,i)} \sim TN\left(\mu_0^{(k^{(g,i)}, s_t^{(g,i)})}, \Sigma_0^{(k^{(g,i)}, s_t^{(g,i)})} \mid b_t^{(g,i)} > l_g\right)$$
    - 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 (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 (24)
  - Sample  $\sigma_t^{2(g,j)}$  based on (25)

End

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

 TABLE I  
 HYPERPARAMETER SPECIFICATION FOR NUMERICAL SIMULATION

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

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

Suppose the failure time is defined as the first time the true degradation signal hits the predefined failure threshold  $\Gamma$ , i.e.,  $\inf\{t : \bar{y}_t > \Gamma\}$ . Denote  $R_t$  as the RUL at the current time  $t$ . Then the distribution of  $R_t$  can be expressed by

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

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 (10). Based on the generated samples, the RUL distribution can be approximated by

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



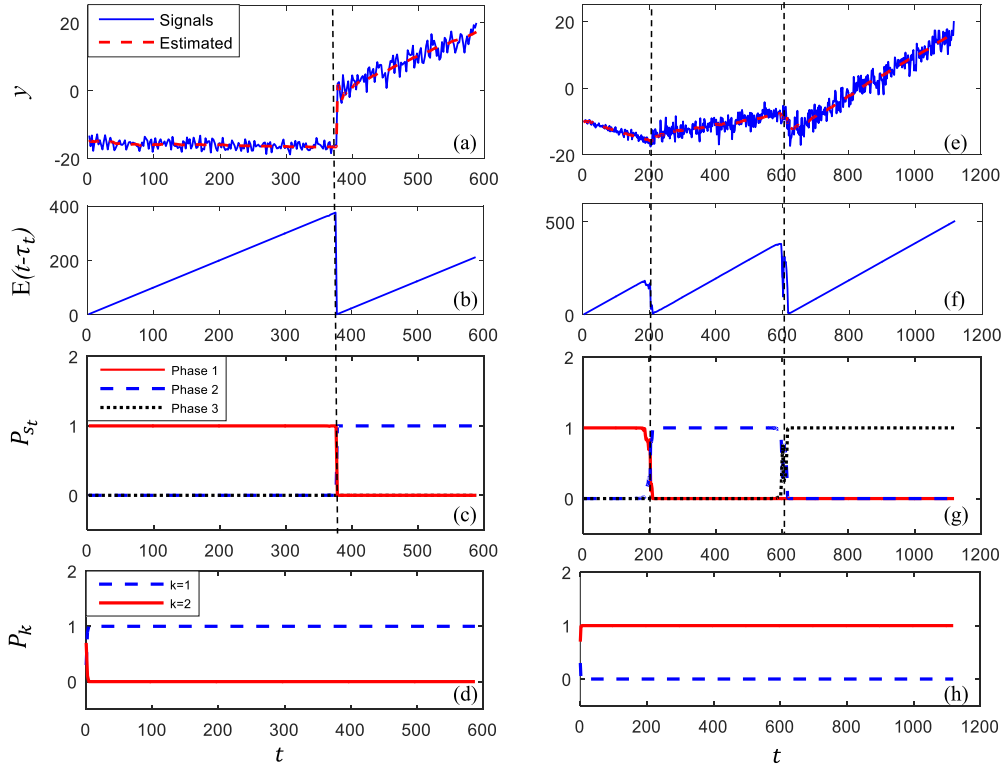


Fig. 3. Illustration of the SPF-based online monitoring of degradation signals with two phases (left panel) and three phases (right panel). (a) and (e): Degradation signals and estimated signals; (b) and (f): The estimated duration of the current linear phase; (c) and (g): The probability mass function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

Conditioning that there are  $k + 1$  line segments or phases for a working unit, the failure will not occur before the  $(k + 1)$ th phase. Therefore, the probability  $P(\bar{y}_{t+j} < \Gamma | \mathbf{x}_{t+j}^{(g,i)})$  can be calculated by

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

where  $1(\cdot)$  is the indicator function.

## 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 section, 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

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

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 hyperparameters 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 hyperparameters are not listed here. In the SPF algorithm, the number of particles for each category is set to be  $N = 5000$ . Fig. 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.

Fig. 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)],

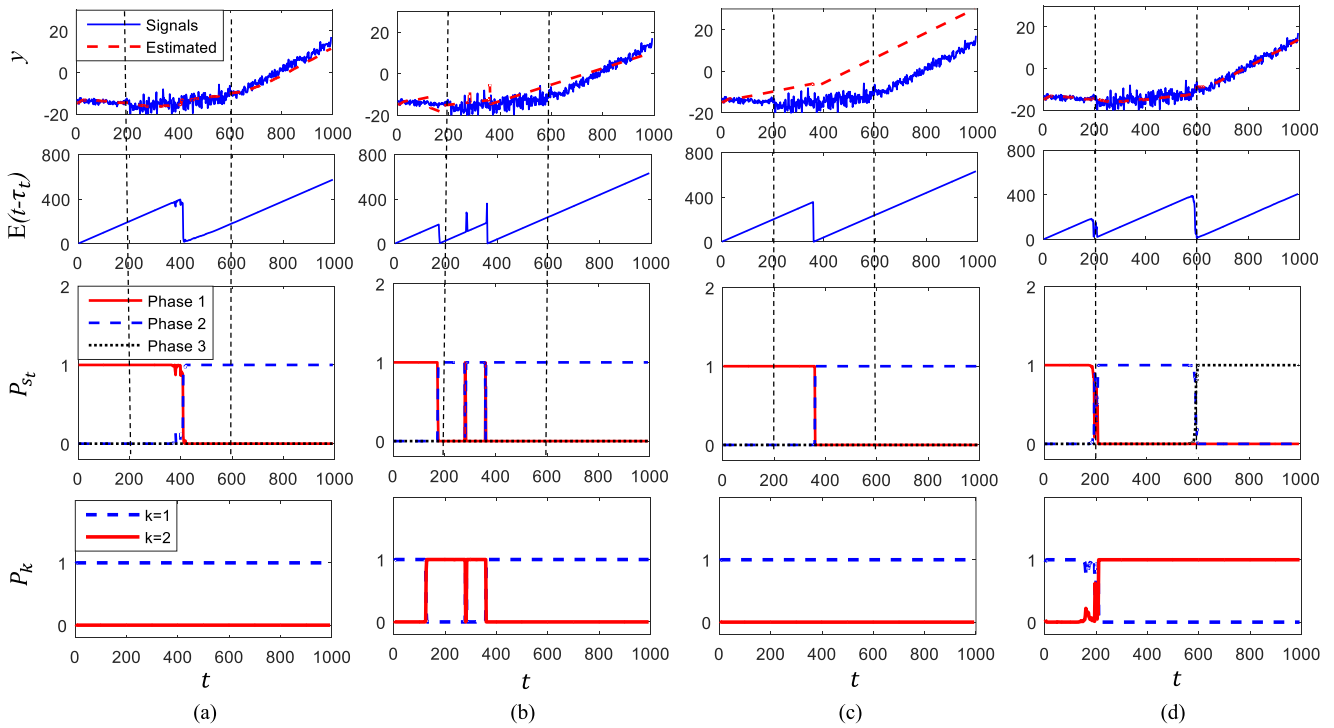


Fig. 4. Comparison of the proposed SPF algorithm to other PF algorithms without either partial Gibbs move or stratified strategy: (a) No stratification, Gibbs move; (b) stratification, no Gibbs move; (c) no stratification, no Gibbs move; and (d) proposed SPF with both particle Gibbs move and stratified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

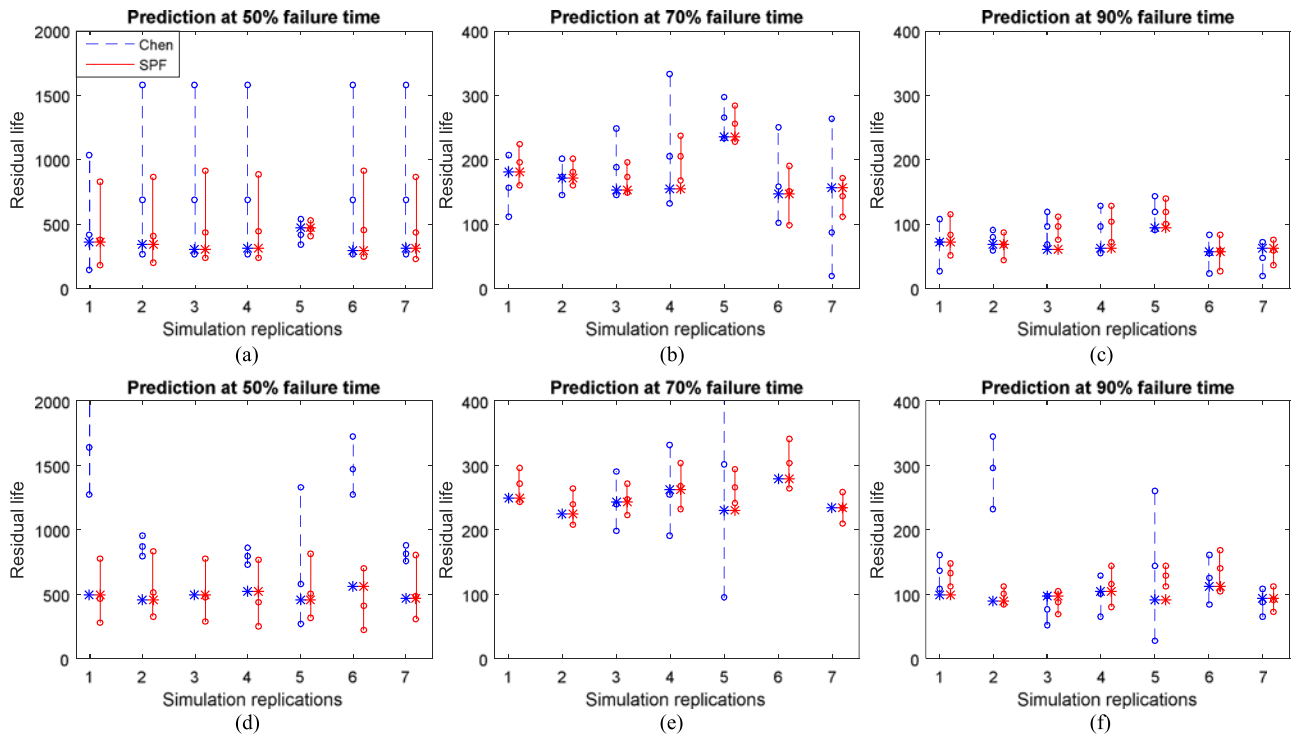


Fig. 5. Prediction intervals of seven simulated CM signals. The  $\circ$  represents the 5%, 50%, 95% quantiles of the RUL distributions,  $*$  is the actual RUL. (a)–(c) Prediction intervals for two-phase signals and (d)–(f) Prediction intervals for three-phase signals.

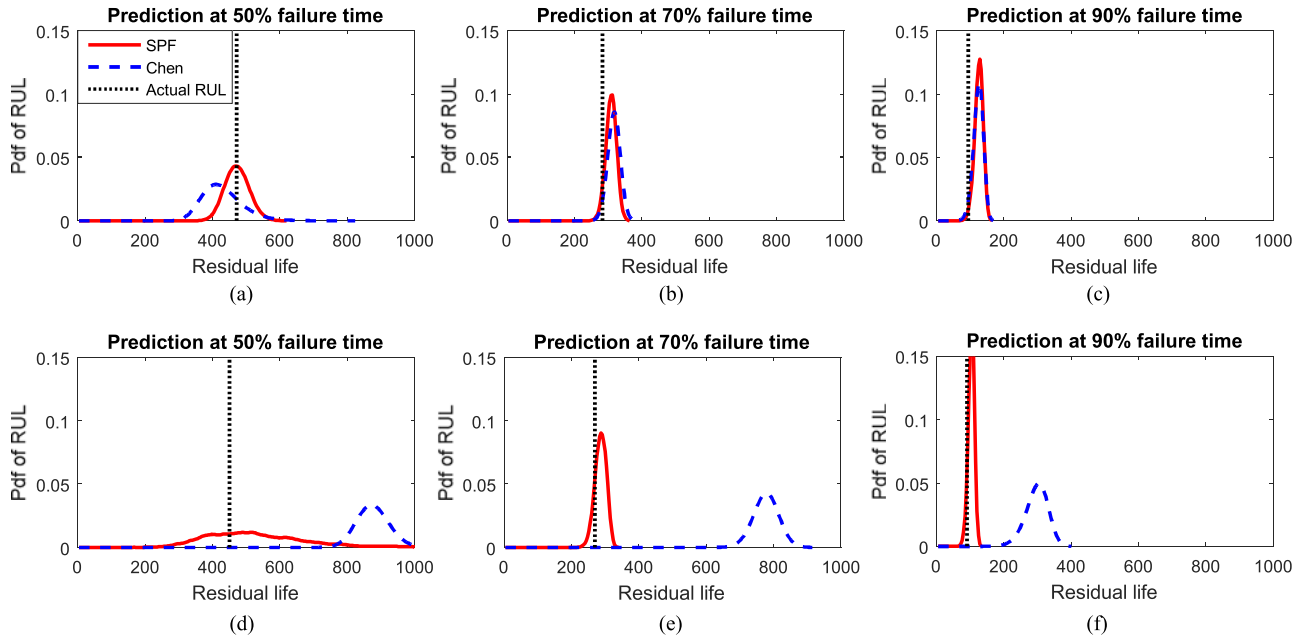


Fig. 6. Comparison of the pdf of the RUL. (a)–(c) two-phase signal and (d)–(f) three-phase signal.

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 [17], 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. Fig. 5 shows the comparison of the prediction intervals of seven simulated two-phase signals and seven three-phase signals predicted at 50%, 70%, and 90% of actual failure time. Fig. 6 shows the detailed RUL prediction of the 5th and the 2nd unit of the seven signals of each category in Fig. 5. From Fig. 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 Fig. 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 Figs. 5(d)–(f) and 6(d)–(f). It is expected since the two-phase model is inadequate to model signals with three phases.

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

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

Method	Updating	Prediction				
		25	50	75	100	125
	300					
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

To evaluate the overall performance, we use the root mean square deviation (RMSD), which is defined as  $\text{RMSD} = \sqrt{E(R - R_{\text{true}})^2}$ , where  $R$  and  $R_{\text{true}}$  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 ten 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

TABLE IV  
ESTIMATED HYPERPARAMETERS OF THE PRIOR DISTRIBUTIONS

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

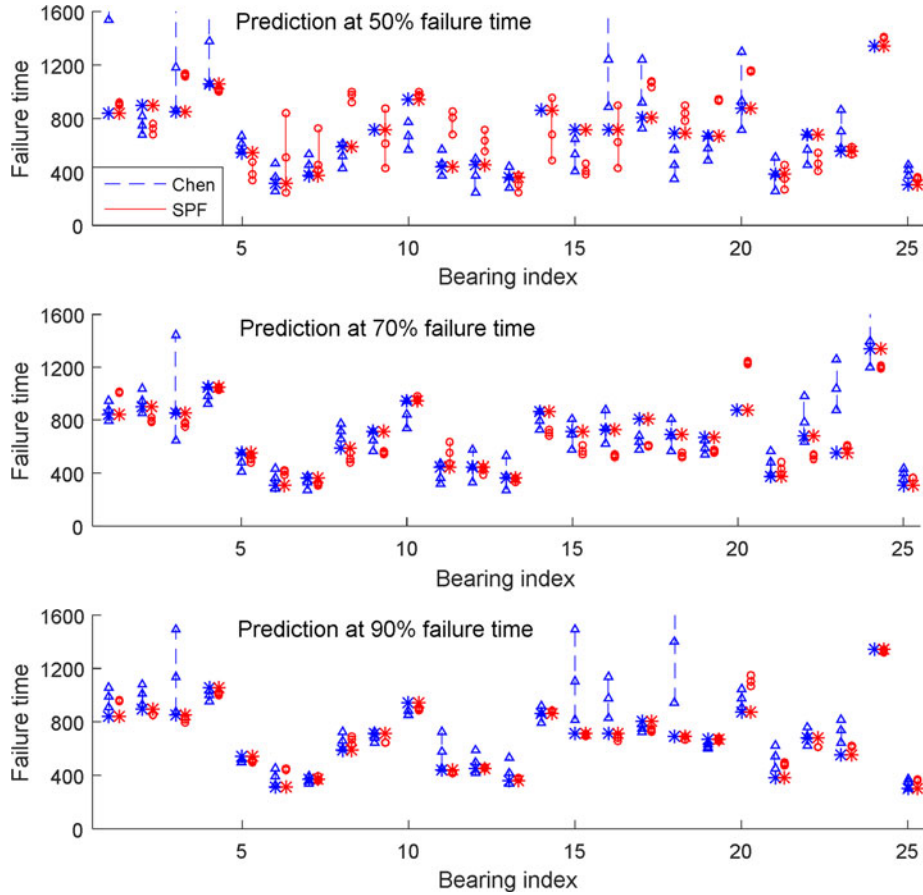


Fig. 7. Prediction intervals of the 25 bearing signals.

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.

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 Fig. 4 we can see that the model updating is quite accurate with only 500 particles.

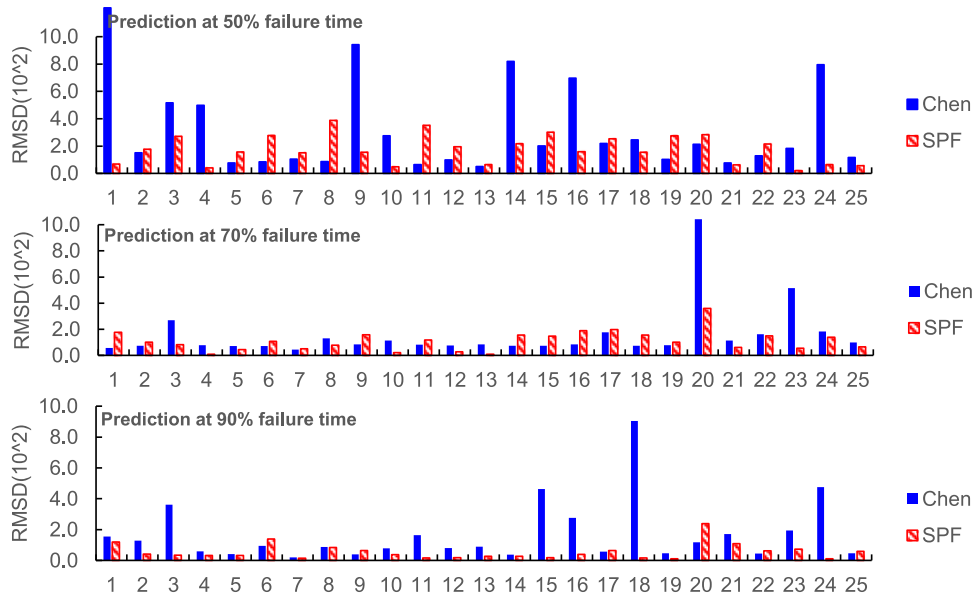


Fig. 8. RMSD of the 25 bearing signals.

### B. Degradation Monitoring of the Rotational Bearings

In this section, the proposed method is applied to the real degradation signals of rotational bearings [6], [12], [17]. 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 min 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 with three phases have the minimum BIC. The estimated hyperparameters 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.

Fig. 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. Fig. 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, the GLLR method [6] and an extension of SPF method with three change-points (SPF-CP3) or four phases 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

TABLE V  
COMPARISON OF THE SPF METHOD WITH OTHER METHODS

Method	RMSD		
	50%	70%	90%
GLLR	356.5	234.2	227.8
Chen	318.4	156.9	169.4
SPF	176.4	110.8	56.9
SPF-CP3	186	112.5	52.4

remaining data are fitted using Bayesian simple linear regression. Clearly, our method outperforms Chen's method and the GLLR method at all three prediction times. The GLLR method has the largest prediction error on this dataset. Comparing SPF and SPF-CP3, we find that adding one more phase will reduce the prediction performance at the early prediction stage (50% and 70% prediction time) while slightly improve the accuracy at 90% prediction time. The reason is that the addition of another phase could introduce extra uncertainty in change-point prediction at early stages, while at the late stage, the model of the last segment of the four-phase model is more accurate in modeling the degradation signal.

## VI. CONCLUSION AND DISCUSSION

In this paper, we propose a multiple-phase modeling of degradation signals for health CM and RUL 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

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

where

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

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

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

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

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

$$(\beta_t | y_{1:t}, \sigma_t^2, \tau_t) \sim TN(\mu_t, \Sigma_t | b_t > l).$$

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

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

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

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

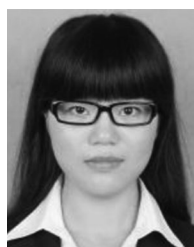
Therefore

$$\begin{aligned} (\sigma_t^2 | \beta_t, y_{1:t}, \tau_t) &\sim IG \left( \alpha_1 + \frac{t-\tau_t}{2}, \alpha_2 \right. \\ &\quad \left. + \frac{\|y_{\tau_t+1:t}^T - \mathbf{X}_{\tau_t+1:t} \beta_t\|^2}{2} \right). \end{aligned}$$

#### REFERENCES

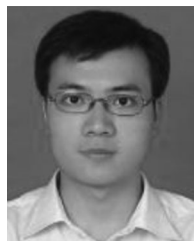
- [1] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, pp. 1483–1510, 2006.
- [2] R. Kothamasu, S. H. Huang, and W. H. VerDuin, "System health monitoring and prognostics—A review of current paradigms and practices," in *Handbook of Maintenance Management and Engineering*. New York, NY, USA: 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 Trans.*, vol. 40, pp. 629–639, 2008.
- [4] M. Pecht and R. Jaai, "A prognostics and health management roadmap for information and electronics-rich systems," *Microelectron. Rel.*, vol. 50, pp. 317–323, 2010.
- [5] C. Okoh, R. Roy, J. Mehnert, 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 Trans.*, vol. 37, pp. 543–557, 2005.
- [7] Z. S. Ye and M. Xie, "Stochastic modelling and analysis of degradation for highly reliable products," *Appl. Stoch. Models Bus. Ind.*, vol. 31, pp. 16–32, 2015.
- [8] Z.-S. Ye, Y. Wang, K.-L. Tsui, and M. Pecht, "Degradation data analysis using Wiener processes with measurement errors," *IEEE Trans. Rel.*, vol. 62, no. 4, pp. 772–780, Dec. 2013.
- [9] 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," *Rel. Eng. Syst. Safety*, vol. 95, pp. 49–57, 2010.
- [10] O. E. Dragomir, R. Gourtveau, N. Zerhouni, and F. Dragomir, "Framework for a distributed and hybrid prognostic system," *IFAC Proc. Volumes*, vol. 40, pp. 431–436, 2007.

- [11] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation—A review on the statistical data driven approaches," *Eur. J. Oper. Res.*, vol. 213, pp. 1–14, 2011.
- [12] N. Gebraeel, "Sensory-updated residual life distributions for components with exponential degradation patterns," *IEEE Trans. Autom. Sci. Eng.*, vol. 3, pp. 382–393, 2006.
- [13] J. Son, Y. Zhang, C. Sankavaram, and S. Zhou, "RUL prediction for individual units based on condition monitoring signals with a change point," *IEEE Trans. Rel.*, vol. 64, no. 1, pp. 182–196, Mar. 2015.
- [14] S. J. Bae and P. H. Kvam, "A nonlinear random-coefficients model for degradation testing," *Technometrics*, vol. 46, pp. 460–469, 2004.
- [15] J. Feng, Q. Sun, and T. Jin, "Storage life prediction for a high-performance capacitor using multi-phase Wiener degradation model," *Commun. Statist.-Simul. Comput.*, vol. 41, pp. 1317–1335, 2012.
- [16] T. S. Ng, "An application of the EM algorithm to degradation modeling," *IEEE Trans. Rel.*, vol. 57, no. 1, pp. 2–13, Mar. 2008.
- [17] N. Chen and K. L. Tsui, "Condition monitoring and remaining useful life prediction using degradation signals: Revisited," *IIE Trans.*, vol. 45, pp. 939–952, 2013.
- [18] S.-T. Tseng, M. Hamada, and C.-H. Chiao, "Using degradation data to improve fluorescent lamp reliability," *J. Qual. Technol.*, vol. 27, pp. 363–369, 1995.
- [19] C. Chiao and M. Hamada, "Using degradation data from an experiment to achieve robust reliability for light emitting diodes," *Qual. Rel. Eng. Int.*, vol. 12, pp. 89–94, 1996.
- [20] S. J. Bae and P. H. Kvam, "A change-point analysis for modeling incomplete burn-in for light displays," *IIE Trans.*, vol. 38, pp. 489–498, 2006.
- [21] Y. Li and P. Nilkitsaranont, "Gas turbine performance prognostic for condition-based maintenance," *Appl. Energy*, vol. 86, pp. 2152–2161, 2009.
- [22] 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 Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6304–6315, Nov. 2014.
- [23] 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 Trans. Cybern.*, vol. 47, no. 1, pp. 67–80, Jan. 2017.
- [24] 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 Trans. Ind. Informat.*, vol. 12, no. 3, pp. 924–932, Jun. 2016.
- [25] Z. Ye, N. Chen, and K. L. Tsui, "A Bayesian approach to condition monitoring with imperfect inspections," *Qual. Rel. Eng. Int.*, vol. 31, pp. 513–522, 2015.
- [26] R. J. Meinhold and N. D. Singpurwalla, "Understanding the Kalman filter," *Amer. Statistician*, vol. 37, pp. 123–127, 1983.
- [27] A. Hannart and P. Naveau, "An improved Bayesian information criterion for multiple change-point models," *Technometrics*, vol. 54, pp. 256–268, 2012.
- [28] 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 Trans. Autom. Sci. Eng.*, vol. 13, no. 2, pp. 688–700, Apr. 2016.
- [29] Y. Hou, J. Wu, and Y. Chen, "Online steady state detection based on rao-blackwellized sequential Monte Carlo," *Qual. Rel. Eng. Int.*, vol. 32, pp. 2667–2683, 2016.
- [30] J. Wu, Y. Chen, and S. Zhou, "Online detection of steady-state operation using a multiple-change-point model and exact Bayesian inference," *IIE Trans.*, vol. 48, pp. 599–613, 2016.
- [31] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, pp. 461–464, 1978.
- [32] R. Killick, P. Fearnhead, and I. Eckley, "Optimal detection of change-points with a linear computational cost," *J. Amer. Statist. Assoc.*, vol. 107, pp. 1590–1598, 2012.
- [33] 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 Trans. Instrum. Meas.*, vol. 62, no. 2, pp. 364–376, Feb. 2013.
- [34] E. Zio and G. Peloni, "Particle filtering prognostic estimation of the remaining useful life of nonlinear components," *Rel. Eng. Syst. Safety*, vol. 96, pp. 403–409, 2011.
- [35] Q. Miao, L. Xie, H. Cui, W. Liang, and M. Pecht, "Remaining useful life prediction of lithium-ion battery with unscented particle filter technique," *Microelectron. Rel.*, vol. 53, pp. 805–810, 2013.
- [36] N. Li, Y. Lei, J. Lin, and S. X. Ding, "An improved exponential model for predicting remaining useful life of rolling element bearings," *IEEE Trans. Ind. Electron.*, vol. 62, no. 12, pp. 7762–7773, Dec. 2015.
- [37] Y. Hu, P. Baraldi, F. Di Maio, and E. Zio, "A particle filtering and kernel smoothing-based approach for new design component prognostics," *Rel. Eng. Syst. Safety*, vol. 134, pp. 19–31, 2015.
- [38] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statist. Comput.*, vol. 10, pp. 197–208, 2000.
- [39] W. R. Gilks and C. Berzuini, "Following a moving target—Monte Carlo inference for dynamic Bayesian models," *J. R. Statist. Soc., Ser. B (Statist. Methodol.)*, vol. 63, pp. 127–146, 2001.
- [40] N. Chopin, "Dynamic detection of change points in long time series," *Ann. Inst. Statist. Math.*, vol. 59, pp. 349–366, 2007.



**Yuxin Wen** received the B.S. degree in medical informatics engineering from Sichuan University, Sichuan, China, in 2011, and the M.S. degree in biomedical engineering from Zhejiang University, Zhejiang, China, in 2014. She is currently working toward the Ph.D. degree in electrical and computer engineering at the University of Texas at El Paso, El Paso, TX, USA.

Her research interests are focused on statistical modeling, prognostics, and reliability analysis.



**Jianguo Wu** 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 the M.S. degree in statistics in 2014 and the Ph.D. degree in industrial and systems engineering in 2015, both from the University of Wisconsin-Madison, Madison, WI, USA.

He is an Assistant Professor in the Department of Industrial, Manufacturing and Systems Engineering, University of Texas at El Paso, El Paso, TX, USA.

He is a member of the Institute for Operations Research and the Management Sciences (INFORMS), the Institute of Industrial and Systems Engineers (IISE), and the Society of Manufacturing Engineers (SME). 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.



**Yuan Yuan** received the B.E. degree in automation from Tsinghua University, Beijing, China, in 2006, and the M.S. degree in industrial and systems engineering in 2010, the M.S. degree in statistics in 2011, and the Ph.D. degree in industrial and systems engineering in 2014 from the University of Wisconsin-Madison, Madison, WI, USA.

She is a Research Scientist at the IBM Research-Singapore. Her research mainly focuses on data analytics, in particular developing innovative and generic data-driven modeling and analysis methodologies for

complex systems with massive data.

Dr. Yuan 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) and the featured article award of IE magazine (2010).