Received March 25, 2019, accepted April 10, 2019, date of publication April 17, 2019, date of current version May 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2911307

An Adaptive Prognostic Approach for Newly Developed System With Three-Source Variability

XI WANG, CHANGHUA HU, XIAOSHENG SI, JIANFEI ZHENG, HONG PEI, YONG YU, AND ZIQIANG REN

Rocket Force University of Engineering, Xi'an 710025, China

Corresponding author: Xiaosheng Si (sxs09@mails.tsinghua.edu.cn)

This work was supported in part by the Natural Science Foundation of China (NSFC) under Grant 61773386, Grant 61833016, Grant 61573365, Grant 61573366, and Grant 61673311, and in part by the Young Elite Scientists Sponsorship Program (YESS) by the China Association for Science and Technology (CAST) under Grant 2016QNRC001.

ABSTRACT Remaining useful life (RUL) estimation is the key of prognostics and health management (PHM) technology and is an effective way to ensure the safe and reliable operation of equipment. Aiming at the lack of historical data and prior information for the newly developed small-sample systems, an adaptive RUL estimation method based on the expectation maximization (EM) algorithm is proposed with three-source variability. First, a degradation model based on a Wiener process is established to incorporate three-source variability and dynamic sampling interval, and the analytical solution of RUL distribution is derived in the sense of the first hitting time. Second, an adaptive parameter estimation method based on the EM algorithm is proposed to update the model parameters by using the condition monitoring (CM) data from one working system running up to the current moment. Finally, a practical example of a gyroscope in an inertial navigation system is provided to substantiate the effectiveness and superiority of the proposed method. The results indicate that the proposed method can efficiently improve the accuracy of the RUL estimation.

INDEX TERMS Remaining useful life, three-source variability, expectation maximization, dynamic sampling interval.

I. INTRODUCTION

With the rapid development of high technology, a large number of new devices have been successfully applied to aerospace and weaponry systems, such as aerospace engines and gyroscopes. Once the failure occurs for these devices, the consequences will often be unimaginable, and thus its reliability and safety put forward higher requirements. Prognostics and health management (PHM) technology has become a hot topic in the field of reliability research in recent years [1]. Its core idea is to estimate the remaining useful life (RUL) of a system by using the condition monitoring (CM) data from the system running up to the current moment, and taking corresponding maintenance measures accordingly to reduce the risk of system failure and improve reliability and security [2]–[4]. However, these newly developed systems, which are vital and small in quantity, generally have problems such as lacking of historical data and prior information during performing the PHM.

The associate editor coordinating the review of this manuscript and approving it for publication was Bora Onat.

As the key of PHM technology, the RUL estimation has received extensive attention from the academic community in recent years. According to the review by [7] and references therein, the current RUL estimation approaches are mainly classified as knowledge-based approaches [8], [9], physical model-based approaches [10], [11], and data-driven approaches [12], [13]. Knowledge-based approaches aim to predict the RUL of the concerned system by modeling an expert system or a fuzzy system. This type of approaches primarily depends on expert's experience with the CM data, and thus it is not suitable for the newly developed systems. Physical model-based approaches rely on modeling the underlying degradation process to be able to predict the occurrence time of failures, but a deep understanding of the concerned system's physics is required. In contrast, data-driven approaches are more widely used in the field of degradation modeling and the RUL estimation due to their flexibility in application. Nowadays, data-driven approaches primarily include machine learning methods [14] and stochastic model-based methods [5]. Machine learning methods usually utilize the CM data to fit the evolution law of the concerned system, and then extrapolate to the failure threshold to achieve the RUL estimation. However, such methods are difficult to obtain the probability distribution which reflects the RUL uncertainty. Contrastively, stochastic model-based methods can establish a degradation model for the CM data and obtain the probability distribution of the RUL based on probability theory, which provide a basis for PHM. Compared with other methods, stochastic model-based methods occupy a dominant position. Si *et al*. [5] systematically reviewed stochastic model-based methods, such as Wiener processes, Gamma processes and Markov chain processes. So far, the Wiener process has been widely used in the field of system reliability analysis and the RUL estimation because of its good mathematical properties and its suitability for describing non-monotonic stochastic degradation process in engineering practice [14]. At present, the Wiener process has been successfully applied to model the stochastic degradation systems, including gyroscopes [26], hard disk drives [13], proton exchange membrane fuel cells [15], rotational bearings [36], and liquid coupling devices [22].

In practice, the degradation processes of systems are usually affected by the multiple sources of variability contributing to the uncertainty of the RUL estimation. Therefore, the multiple sources of variability should be incorporated into degradation modelling to improve the accuracy of the RUL estimation. According to existing studies [6], [13], [16]–[19], temporal variability, unit-to-unit variability, and measurement variability widely coexist in the degradation process of a system. Temporal variability indicates the inherent uncertainty associated with the degradation process over time, which is the reason for modeling with a stochastic process [20]. Unit-to-unit variability depicts the heterogeneity among different degradation units [19]. Specifically, the degradation rate of units is practically different from each other though the shapes of the degradation paths are similar. Measurement variability means the difference between the actual degradation state and the measurements. In practice, due to the effect of external noise, disturbance, the precision of instrument, etc., some measurement errors may inevitably be introduced during the observation process called imperfect measurements [13]. Therefore, the CM data can only partially reflect the actual degradation state of the system.

From the above introduction, it has been recognized that an appropriate stochastic degradation model should incorporate three-source variability simultaneously. However, most published works using stochastic degradation models focused on the RUL estimation with one or two sources variability [15], [20], [22], [23], [26]–[28], [36], [38].

Recently, for degradation modeling with three-source variability, Ye *et al*. [13] established a Wiener process with mixed-effects considering three-source variability, which could be used to fit the hard disk drive wear data, but it is limited to modeling the degradation process only. Si *et al.* [6] proposed a linear Wiener degradation process with three-source variability and incorporated the effect of three-source variability into the RUL estimation. In this work, the RUL estimation results considering three-source variability were compared with the results considering only one or two sources variability by the degradation data of gyroscopes. The results indicated that considering three-source variability could significantly improve the goodness of fit of the model and the accuracy of the RUL estimation. On this basis, Zheng *et al*. [19] further verified the superiority of considering three-source variability through a simulation study and the degradation data of 2017-T4 aluminum alloy. To sum up, it is necessary to consider three-source variability simultaneously in order to obtain more accurate estimation results within the framework of stochastic modeling. However, a common assumption in these works is that there are multiple sets of historical data for degrading systems. However, for the newly developed small-sample system, the above methods are not applicable and the related research is very limited. This motivates our research in this paper, i.e. developing an adaptive prognostic approach for newly developed system with three-source variability.

In this paper, for the newly developed small-sample systems, we present an adaptive prognostic approach based on the EM algorithm for RUL estimation with three-source variability. Firstly, we establish a Wiener process based state-space model with three-source variability. Then, the analytical RUL distribution with three-source variability is formulated through the joint posterior estimation of the model's random parameter and the underlying degradation state. Based on the CM data of one working system running to the current time, a parameter estimation method based on the Rauch-Tung-Striebel (RTS) algorithm and EM algorithm is proposed to realize adaptive estimation and online updating of model parameters. Finally, we provide a practical example of a gyroscope in an inertial navigation system to verify the effectiveness of the proposed method. The results indicate that the proposed method is not only suitable for the newly developed small-sample systems, but also significantly improves the accuracy of the RUL estimation.

The remainder of this paper is organized as follow: In Section II, the recently related works are reviewed and discussed. Section III gives the degradation modeling for RUL estimation with three-source variability. An adaptive parameter estimation method is proposed in Section IV. Section V provides a practical case study of a gyroscope for demonstration. The paper is concluded in Section VI. All the lengthy proofs in this paper are provided in the Appendix.

II. RELATED WORKS

There have been rapid development and extensive application based on degradation modeling for RUL estimation, such as in the examples in [15], [20]–[23], [26]–[28]. However, it should be noted that the above works focused mostly on the RUL estimation with one or two sources variability. For example, in [15], the authors considered the temporal variability and unit-to-unit variability to estimate the RUL of the system, and utilized the EM algorithm to identify the unknown parameters. However, the impact of measurement variability on the RUL estimation was ignored. In addition, the work in [27] proposed a degradation model based on a Wiener process with Kalman filtering technique and Bayesian updating to estimate the RUL of the system, but only considered the temporal variability and unit-to-unit variability in the RUL estimation as well. Therefore, many researchers have begun to study the RUL estimation problems with three-source variability.

However, in these works, most existing methods for the RUL estimation either assume the existence of historical degradation data for multiple sets of similar systems or determine the model parameters based on subjective information such as expert knowledge and historical experience in model parameter estimation [6], [13], [17]–[23]. For the newly developed small-sample system, the existing RUL estimation methods with three-source variability are no longer applicable due to the lack of such historical data. For example, Si *et al.* [6] addressed the RUL estimation problem with three-source variability based on a Wiener process. Following this work, Zhang *et al.* [18] and Zheng *et al.* [19] considered the stochastic degradation modeling problems under the three-source variability. However, the common shared be these studies is to assume the existence of historical degradation data for multiple sets of similar systems in parameter estimation, and then the hyper-parameters in the models are determined by off-line estimation method. Although the random parameters can be adaptively estimated with the real-time measurement data of the system, once the hyper-parameters are determined, they will not be updated. However, when the parameter estimation is not accurate enough, the RUL estimation result will be affected and it is even difficult to accurately predict. Furthermore, Wang and Tsui [20] predicted the RUL of lithium batteries, but the estimation of hyper-parameters was still based on several batteries' historical degradation data of the same type. To solve this problem, many researchers have conducted the related research. For example, a regression model was proposed in [24] to describe the degradation process, and Bayesian updating and EM algorithm were used for parameter estimation. However, the regression model is difficult to describe the stochastic characteristics of the degradation process, and the RUL estimation problem with three-source variability is not considered. In addition, Alamaniotis *et al.* [8] modeled the degradation system with expert knowledge and historical experience to compensate for the lack of historical data, but the model depended on the expert's experience with the degradation system. Consequently, the study of the RUL estimation using degradation models with three-source variability for small-sample systems is still very limited. On the other hand, most published works default to periodic monitoring devices to obtain the CM data. However, due to some subjective or objective reasons in practice, such as changes in the testing plan and experimental equipment failures, the CM cannot always be carried out in strict accordance with the planned cycle. Therefore, a more general form should be considered, i.e., the sampling interval for the CM is dynamic. So far, there is still very scarce research on this issue.

III. DEGRADATION MODELING FOR RUL ESTIMATION WITH THREE-SOURCE VARIABILITY

In this paper, the linear stochastic degradation process $\{X(t), t \geq 0\}$ is modeled by a Wiener process where $X(t)$ is driven by a standard Brownian motion (SBM) $B(t)$. Then, $X(t)$ can be represented as

$$X(t) = X(0) + \lambda t + \sigma B(t) \tag{1}$$

where λ is the drift coefficient that is a random parameter used to represent the degradation rate of the system, σ is the diffusion coefficient which is a constant used to represent the common features of all degraded systems. $B(t)$ is the SBM with $\sigma B(t) \sim N(0, \sigma^2 t)$ for $t > 0$, representing the dynamic characteristics of the underlying degradation process. Without loss of generality, it is assumed that the initial state $X(0) = 0$ in the following.

In fact, each degraded system may experience different experimental conditions during normal operation, and thus different degraded systems may have different degradation paths. As a result, it is necessary to consider the unit-to-unit variability in the degradation model. In this paper, the random parameter λ is used to describe the unit-to-unit variability, and the SBM $B(t)$ is used to describe the temporal variability. It is further assumed that $\lambda \sim N(\mu_\lambda, \sigma_\lambda^2)$, and is independent of $\{B(t), t > 0\}$. All of the above assumptions are reasonable and are widely used in the field of degradation modeling and the RUL estimation [5], [6], [18]–[20].

In addition, due to the effect of external noise, disturbance, the precision of instrument, etc., the CM data inevitably has measurement error in engineering practice. To describe this measurement variability, the degradation measurement process $\{Y(t), t > 0\}$ can be mathematically formulated as

$$Y(t) = X(t) + \varepsilon \tag{2}$$

where ε is an Gaussian observation noise with $\varepsilon \sim N(0, \gamma^2)$, representing the random measurement error. Furthermore, we assume that ε, λ and $B(t)$ are mutually independent. By now, three-source variability is incorporated into the degradation modeling process.

To integrate dynamic sampling interval into the degradation model and achieve the adaptive estimation for RUL under three-source variability, we consider establishing an updating mechanism for the drift coefficient λ by a model $\lambda_k = \lambda_{k-1} + \alpha$ at the current time t_k, where $\alpha \sim N(0, \nu^2)$ and the initial coefficient λ_0 follows $\lambda_0 \sim N(\mu_\lambda, \sigma_\lambda^2)$. Therefore, based on (1) and (2), suppose the degradation process and measurement process are discretized at the discrete time point $t_k (k = 1, 2, \ldots)$. For convenience, let $\mathbf{Y}_{1:k} = \{y_1, y_2, \ldots, y_k\}$ represent the CM data set and $y_k = Y(t_k)$ denote the CM data at t_k. Similarly, let $\mathbf{X}_{1:k} = \{x_1, x_2, \ldots, x_k\}$ represent the underlying degradation states set and $x_k = X(t_k)$. From (2), We further express the measurement equation as $y_k = x_k + \varepsilon_k$, where the random measurement errors ε_k are assumed to be independent and identically distributed (i.i.d.). As such, the degradation equations with three-source variability can be constructed within the framework of state-space modeling as

$$\begin{cases} x_k = x_{k-1} + \lambda_{k-1}(t_k - t_{k-1}) + \zeta_k \\ \lambda_k = \lambda_{k-1} + \alpha \\ y_k = x_k + \varepsilon_k \end{cases} \tag{3}$$

where $\zeta_k = \sigma[B(t_k) - B(t_{k-1})]$ and $\varepsilon_k \sim N(0, \gamma^2)$. According to the property of the SBM, we further have $\zeta_k \sim N\left[0, \sigma^2(t_k - t_{k-1})\right]$. It is worth noting that in order to distinguish the existing periodic sampling, the dynamic sampling interval of the CM is described by $t_k - t_{k-1}$, and thus the periodic sampling is included as a special case. Specifically, the periodic sampling means that $t_k - t_{k-1}$ is a constant. Here, this paper considers $t_k - t_{k-1}$ as a variable which propagates its impact on the derivation process of the parameter estimation.

Based on other degradation modeling works such as [5], [20], and [25], a system's lifetime is defined as the time when the underlying degradation state $\{X(t), t \geq 0\}$ reaches a preset failure threshold ω for the first time, i.e., the first hitting time (FHT). Therefore, a system's RUL is defined as the effective time interval from the current time to the system failure time. According to the concept of the FHT, the RUL L_k of a system at the current time t_k can be defined as

$$L_k = \inf\{l_k > 0 : X(l_k + t_k) \geq \omega\} \tag{4}$$

with the conditional probability density function (PDF) $f_{L_k|\mathbf{Y}_{1:k}}(l_k|\mathbf{Y}_{1:k})$ where $\mathbf{Y}_{1:k}$ is all the CM data up to t_k.

In the following, the focus is on solving conditional PDF $f_{L_k|\mathbf{Y}_{1:k}}(l_k|\mathbf{Y}_{1:k})$ of the RUL based on $\mathbf{Y}_{1:k}$ and making it continuously updated with the accumulation of the CM data. According to the established degradation model (3), We regard the underlying degradation state x_k and the drift coefficient λ as hidden states. Then, the state-space model (3) can be further represented as

$$\begin{cases} \mathbf{z}_k = \mathbf{A}_k \mathbf{z}_{k-1} + \boldsymbol{\eta}_k \\ y_k = \mathbf{C}\mathbf{z}_k + \varepsilon_k \end{cases} \tag{5}$$

where $\mathbf{z}_k \in R^{2\times 1}$, $\mathbf{A}_k \in R^{2\times 2}$, $\boldsymbol{\eta}_k \in R^{2\times 1}$, $\mathbf{C} \in R^{1\times 2}$, and $\boldsymbol{\eta}_k \sim N(0, \mathbf{Q}_k)$, with

$$\mathbf{z}_k = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix}, \quad \mathbf{A}_k = \begin{bmatrix} 1 & t_k - t_{k-1} \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{\eta}_k = \begin{bmatrix} \zeta_k \\ \alpha \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\mathrm{T}}, \quad \mathbf{Q}_k = \begin{bmatrix} \sigma^2(t_k - t_{k-1}) & 0 \\ 0 & \nu^2 \end{bmatrix},$$

respectively. Here, $(\cdot)^{\mathrm{T}}$ denotes the vector transposition.

For convenience, the expectation and variance of $\mathbf{z}_k$, which are estimated on $\mathbf{Y}_{1:k}$, can be defined as

$$\hat{\mathbf{z}}_{k|k} = \begin{bmatrix} \hat{x}_{k|k} \\ \hat{\lambda}_{k|k} \end{bmatrix} = \mathbb{E}(\mathbf{z}_k|\mathbf{Y}_{1:k})$$

$$\mathbf{P}_{k|k} = \begin{bmatrix} \rho^2_{x,k} & \rho^2_{x\lambda,k} \\ \rho^2_{x\lambda,k} & \rho^2_{\lambda,k} \end{bmatrix} = \mathrm{cov}(\mathbf{z}_k|\mathbf{Y}_{1:k}) \tag{6}$$

where

$$\hat{x}_{k|k} = \mathbb{E}(x_k|\mathbf{Y}_{1:k}), \quad \hat{\lambda}_{k|k} = \mathbb{E}(\lambda_k|\mathbf{Y}_{1:k}),$$
$$\rho^2_{x,k} = \mathrm{var}(x_k|\mathbf{Y}_{1:k})$$
$$\rho^2_{\lambda,k} = \mathrm{var}(\lambda_k|\mathbf{Y}_{1:k}), \quad \rho^2_{x\lambda,k} = \mathrm{cov}(x_k\lambda_k|\mathbf{Y}_{1:k})$$

Similarly, the one-step ahead prediction of the expectation and variance can be defined as

$$\hat{\mathbf{z}}_{k|k-1} = \begin{bmatrix} \hat{x}_{k|k-1} \\ \hat{\lambda}_{k|k-1} \end{bmatrix} = \mathbb{E}(\mathbf{z}_k|\mathbf{Y}_{1:k-1})$$

$$\mathbf{P}_{k|k-1} = \begin{bmatrix} \rho^2_{x,k|k-1} & \rho^2_{x\lambda,k|k-1} \\ \rho^2_{x\lambda,k|k-1} & \rho^2_{\lambda,k|k-1} \end{bmatrix} = \mathrm{cov}(\mathbf{z}_k|\mathbf{Y}_{1:k-1}) \tag{7}$$

Based on the above definition, we can utilize the Kalman filtering algorithm to iteratively estimate the joint posterior distribution of the underlying degradation state x_k and the drift coefficient λ at the current time t_k when a new measurement y_k is available. The Kalman filtering algorithm contains two recursive phases as follows.

State estimation:

$$\mathbf{P}_{k|k-1} = \mathbf{A}_k \mathbf{P}_{k-1|k-1}\mathbf{A}_k^{\mathrm{T}} + \mathbf{Q}_k$$
$$\mathbf{K}(k) = \mathbf{P}_{k|k-1}\mathbf{C}^{\mathrm{T}}\left(\mathbf{C}\mathbf{P}_{k|k-1}\mathbf{C}^{\mathrm{T}} + \gamma^2\right)^{-1}$$
$$\hat{\mathbf{z}}_{k|k-1} = \mathbf{A}_k \hat{\mathbf{z}}_{k-1|k-1}$$
$$\hat{\mathbf{z}}_{k|k} = \hat{\mathbf{z}}_{k|k-1} + \mathbf{K}(k)\left(y_k - \mathbf{C}\hat{\mathbf{z}}_{k|k-1}\right)$$

Variance update:

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}(k)\mathbf{C}\mathbf{P}_{k|k-1}$$

where the initial iterative values are specified as

$$\hat{\mathbf{z}}_{0|0} = \begin{bmatrix} 0 \\ \mu_\lambda \end{bmatrix}, \quad \mathbf{P}_{0|0} = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_\lambda^2 \end{bmatrix}.$$

It is worth noting that, the sampling interval is integrated into $\hat{\mathbf{z}}_{k|k}$ and $\mathbf{P}_{k|k}$ by the Kalman filtering algorithm. σ_x^2 is the variance of the initial underlying degradation state x_0. According to the linear Gaussian properties of the Kalman filtering algorithm, the PDF of $\mathbf{z}_k$ based on $\mathbf{Y}_{1:k}$ is bivariate normal distribution with $\mathbf{z}_k \sim N\left(\hat{\mathbf{z}}_{k|k}, \mathbf{P}_{k|k}\right)$. Hence, we can get the initial state $\mathbf{z}_0 \sim N\left(\boldsymbol{\mu}_0, \mathbf{P}_0\right)$, where $\boldsymbol{\mu}_0 = \hat{\mathbf{z}}_{0|0}$ and $\mathbf{P}_0 = \mathbf{P}_{0|0}$.

According to the estimated joint PDF of $\mathbf{z}_k$ and the definition of the RUL in (4), the conditional PDF and the expectation of the RUL can be respectively calculated as follows (8) and (9)

$$f_{L_k|\mathbf{Y}_{1:k}}(l_k|\mathbf{Y}_{1:k}) = \frac{\mathcal{B}_k\left(\mathcal{D}_k^2\rho^2_{\lambda,k} + \mathcal{F}_k\right) - \mathcal{A}_k\mathcal{C}_k\mathcal{D}_k\rho^2_{\lambda,k} - \mathcal{C}_k\mathcal{F}_k\hat{\lambda}_{k|k}}{\mathcal{F}_k\sqrt{2\pi\left(\mathcal{D}_k^2\rho^2_{\lambda,k} + \mathcal{F}_k\right)^3}} \times \exp\left[-\frac{\left(w - \hat{x}_{k|k} - \hat{\lambda}_{k|k}l_k\right)^2}{2\left(\mathcal{D}_k^2\rho^2_{\lambda,k} + \mathcal{F}_k\right)}\right] \tag{8}$$

$$\mathbb{E}(L_k|\mathbf{Y}_{1:k}) = \mathcal{A}_k\mathcal{G}_k - \varphi \tag{9}$$

where

$$\mathcal{A}_k = w - \hat{x}_{k|k} + \varphi\hat{\lambda}_{k|k}, \quad \mathcal{B}_k = \left(w - \hat{x}_{k|k} + \varphi\hat{\lambda}_{k|k}\right)\sigma^2$$

$$\mathcal{C}_k = \left(\sigma^2 + \rho_{x\lambda,k}^2\right)\varphi - \rho_{x,k}^2, \quad \mathcal{D}_k = \varphi + l_k, \ \varphi = \frac{\rho_{x\lambda,k}^2}{\rho_{\lambda,k}^2}$$

$$\mathcal{F}_k = \rho_{x,k}^2 - \rho_{x\lambda,k}^2\varphi + \sigma^2 l_k, \quad \mathcal{G}_k = \frac{\sqrt{2}}{\rho_{\lambda,k}} D\left(\frac{\hat{\lambda}_{k|k}}{\sqrt{2}\rho_{\lambda,k}}\right)$$

The detailed proof for (8) and (9) can be found in [6] and thus is omitted here. When (8) and (9) are used for real-time estimation, the initial parameters of the model including μ_λ, σ_λ^2, σ_x^2, σ^2, ν^2, and γ^2 are unknown. Typically, historical degradation data for multiple sets of similar systems are used to determine the unknown parameters [6], [13], [17]–[23]. In addition, once these parameters are determined by historical degradation data, they are no longer updated in real-time with measurement data. However, for the newly developed small-sample system, due to the lack of historical degradation data and prior information, it is necessary to continuously update all model parameters with the real-time monitored degradation data so that the estimated RUL can better reflect the current health status of the system. For this purpose, we develop an adaptive parameter estimation algorithm based on the EM algorithm in the following section.

IV. ADAPTIVE PARAMETER ESTIMATION

The parameter estimation method presented in this paper can use the CM data $\mathbf{Y}_{1:k}$ of one working system to iteratively update the unknown parameters of the state-space model (5), so that the RUL estimation results do not depend on the selection of the initial parameters, thereby achieving a more accurate RUL estimation. To do so, we first let $\Theta = \left[\boldsymbol{\mu}_0^{\mathrm{T}}, vec\{\mathbf{P}_0\}^{\mathrm{T}}, \sigma^2, \nu^2, \gamma^2\right]^{\mathrm{T}}$ denote the unknown parameter vector, where

$$\boldsymbol{\mu}_0 = \begin{bmatrix} 0 \\ \mu_\lambda \end{bmatrix} \text{ and } \mathbf{P}_0 = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_\lambda^2 \end{bmatrix}, \tag{10}$$

and vec $\{\cdot\}$ operator is a linear transformation which converts the matrix into a column vector.

Then, the log-likelihood function of the CM data $\mathbf{Y}_{1:k}$ with respect to the unknown parameter vector Θ at time t_k is

$$L_k(\Theta) = \log p(\mathbf{Y}_{1:k}|\Theta) \tag{11}$$

where $p(\mathbf{Y}_{1:k}|\Theta)$ is a joint PDF of the CM data $\mathbf{Y}_{1:k}$. The maximum likelihood estimate (MLE) $\hat{\Theta}_k$ of Θ on the basis of $\mathbf{Y}_{1:k}$ can be obtained by maximizing the likelihood function (11). That is

$$\hat{\Theta}_k = \arg\max_{\Theta} L_k(\Theta) \tag{12}$$

However, since we regard $\mathbf{z}_k$ as a hidden variable in (5), then (12) cannot be directly maximized. The EM algorithm provides a feasible solution to solve this problem. The basic principle is to approximate the maximum likelihood estimation of the parameters by maximizing the joint likelihood function $p(\mathbf{z}_k, \mathbf{Y}_{1:k}|\Theta)$, so that the estimated parameter vector Θ can be realized by iterating in the following two steps.

(1) *E-Step*

$$\ell\left(\Theta|\hat{\Theta}_k^{(i)}\right) = \mathbb{E}_{\mathbf{z}_k|\mathbf{Y}_{1:k},\hat{\Theta}_k^{(i)}}\{\log p(\mathbf{z}_k, \mathbf{Y}_{1:k}|\Theta)\} \tag{13}$$

where $\hat{\Theta}_k^{(i)}$ represents the result of parameters iteration in the ith step based on $\mathbf{Y}_{1:k}$.

(2) *M-Step*

$$\hat{\Theta}_k^{(i+1)} = \arg\max_{\Theta}\left\{\ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)\right\} \tag{14}$$

According to the EM algorithm [29], the iterative process starts from the estimated value $\hat{\Theta}_k^{(i)}$ of the ith step in the maximum likelihood sense, and is updated to a better estimated $\hat{\Theta}_k^{(i+1)}$ in the $(i+1)$th step, i.e., as the number of iterations increases, the results of parameter estimation are getting better and better. In the practical application of the EM algorithm, it is generally difficult to obtain a satisfactory estimation value in one iteration. Therefore, it is necessary to perform multiple iterations until a given convergence criterion is satisfied.

The EM algorithm is used to estimate the unknown parameter vector Θ. The complete joint likelihood function for the concerned model (5) is

$$\begin{aligned}
&\log p(\mathbf{z}_k, \mathbf{Y}_{1:k}|\Theta) \\
&= \log p(\mathbf{Y}_{1:k}|\mathbf{z}_k, \Theta) + \log p(\mathbf{z}_k|\Theta) \\
&= -\frac{1}{2}\log|\mathbf{P}_0| - \frac{1}{2}(\mathbf{z}_0 - \boldsymbol{\mu}_0)^{\mathrm{T}}\mathbf{P}_0^{-1}(\mathbf{z}_0 - \boldsymbol{\mu}_0) \\
&\quad - \frac{1}{2}\sum_{j=1}^{k}\left[\log|\mathbf{Q}_j| + (\mathbf{z}_j - \mathbf{A}_j\mathbf{z}_{j-1})^{\mathrm{T}}\mathbf{Q}_j^{-1}(\mathbf{z}_j - \mathbf{A}_j\mathbf{z}_{j-1})\right] \\
&\quad - \frac{1}{2}k\log\gamma^2 - \frac{1}{2}\sum_{j=1}^{k}\frac{(y_j - \mathbf{C}\mathbf{z}_j)^{\mathrm{T}}(y_j - \mathbf{C}\mathbf{z}_j)}{\gamma^2}
\end{aligned} \tag{15}$$

where the matrices $\mathbf{P}_0$ and $\mathbf{Q}_j$ are positive definite matrices.

It is worth noting here that differing from the traditional EM algorithm [30]–[33], the matrices $\mathbf{A}_j$ and $\mathbf{Q}_j$ are varied with the sampling interval because of considering the dynamic sampling interval problem in this paper. We further use E-Step to calculate (15), and obtain the following results.

$$\begin{aligned}
&\ell\left(\Theta|\hat{\Theta}_k^{(i)}\right) \\
&\quad = \mathbb{E}_{\mathbf{z}_k|\mathbf{Y}_{1:k},\hat{\Theta}_k^{(i)}}\{\log p(\mathbf{z}_k, \mathbf{Y}_{1:k}|\Theta)\} \\
&\quad = -\frac{1}{2}\log|\mathbf{P}_0| - \frac{1}{2}\mathrm{Tr}\left\{\mathbf{P}_0^{-1}\mathbb{E}_{\mathbf{z}_k|\mathbf{Y}_{1:k},\hat{\Theta}_k^{(i)}}\right. \\
&\qquad \left.\times\left[(\mathbf{z}_0 - \boldsymbol{\mu}_0)(\mathbf{z}_0 - \boldsymbol{\mu}_0)^{\mathrm{T}}\right]\right\} \\
&\qquad - \frac{1}{2}\sum_{j=1}^{k}\left\{\log|\mathbf{Q}_j| + \mathrm{Tr}\left\{\mathbf{Q}_j^{-1}\left[\Phi - \Psi\mathbf{A}_j^{\mathrm{T}}\right.\right.\right. \\
&\qquad \left.\left.\left. - \mathbf{A}_j\Psi^{\mathrm{T}} + \mathbf{A}_j\Sigma\mathbf{A}_j^{\mathrm{T}}\right]\right\}\right\} \\
&\qquad - \frac{1}{2}k\log\gamma^2
\end{aligned}$$

$$-\frac{1}{2}\mathrm{Tr}\left\{\gamma^{-2}\sum_{j=1}^{k}\left[\left(y_j-\mathbf{C}\mathbf{z}_{j|k}\right)\left(y_j-\mathbf{C}\mathbf{z}_{j|k}\right)^{\mathrm{T}}+\mathbf{C}\mathbf{P}_{j|k}\mathbf{C}^{\mathrm{T}}\right]\right\} \tag{16}$$

where

$$\begin{aligned}\Phi &= \mathbb{E}_{\mathbf{z}_k|\mathbf{Y}_{1:k},\hat{\Theta}_k^{(i)}}\left(\mathbf{z}_j\mathbf{z}_j^{\mathrm{T}}\right)\\ \Psi &= \mathbb{E}_{\mathbf{z}_k|\mathbf{Y}_{1:k},\hat{\Theta}_k^{(i)}}\left(\mathbf{z}_j\mathbf{z}_{j-1}^{\mathrm{T}}\right)\\ \Sigma &= \mathbb{E}_{\mathbf{z}_k|\mathbf{Y}_{1:k},\hat{\Theta}_k^{(i)}}\left(\mathbf{z}_{j-1}\mathbf{z}_{j-1}^{\mathrm{T}}\right)\end{aligned} \tag{17}$$

Obviously, if we want to calculate (16), we must get the conditional expectations given by (17). In this paper, we use the RTS algorithm [34] to compute them and the specific process is as follows.

Step 1: Backwards iteration

$$\begin{aligned}\mathbf{S}_j &= \mathbf{P}_{j|j}\mathbf{A}_{j+1}^{\mathrm{T}}\mathbf{P}_{j+1|j}^{-1}\\ \hat{\mathbf{z}}_{j|k} &= \hat{\mathbf{z}}_{j|j}+\mathbf{S}_j\left(\hat{\mathbf{z}}_{j+1|k}-\hat{\mathbf{z}}_{j+1|j}\right)\\ \mathbf{P}_{j|k} &= \mathbf{P}_{j|j}+\mathbf{S}_j\left(\mathbf{P}_{j+1|k}-\mathbf{P}_{j+1|j}\right)\mathbf{S}_j^{\mathrm{T}}\end{aligned} \tag{18}$$

Step 2: Initialization

$$\mathbf{M}_{k|k} = \left(\mathbf{I}-\mathbf{K}_k\mathbf{C}\right)\mathbf{A}_k\mathbf{P}_{k-1|k-1} \tag{19}$$

Step 3: Backward iteration of covariance

$$\mathbf{M}_{j|k} = \mathbf{P}_{j|j}\mathbf{S}_{j-1}^{\mathrm{T}}+\mathbf{S}_j\left(\mathbf{M}_{j+1|k}-\mathbf{A}_{j+1}\mathbf{P}_{j|j}\right)\mathbf{S}_{j-1}^{\mathrm{T}} \tag{20}$$

where $\mathbf{M}_{j|k} = \mathrm{cov}\left(\mathbf{z}_j, \mathbf{z}_{j-1}|\mathbf{Y}_{1:k}\right)$. $\hat{\mathbf{z}}_{j|j}$, $\mathbf{P}_{j|j}$ and $\mathbf{P}_{j+1|j}$ can be pre-computed by Kalman filter and the following lemma is given for calculating the conditional expectation. Detailed derivation process can be found in [34].

Lemma 1: Given the unknown parameter vector $\hat{\Theta}_k^{(i)}$ estimated at the current time and the state-space model (5), we have

$$\begin{aligned}\mathbb{E}_{\mathbf{z}_k|\mathbf{Y}_{1:k},\hat{\Theta}_k^{(i)}}\left(y_j\mathbf{z}_j^{\mathrm{T}}\right) &= y_j\hat{\mathbf{z}}_{j|k}^{\mathrm{T}}\\ \mathbb{E}_{\mathbf{z}_k|\mathbf{Y}_{1:k},\hat{\Theta}_k^{(i)}}\left(\mathbf{z}_j\mathbf{z}_j^{\mathrm{T}}\right) &= \hat{\mathbf{z}}_{j|k}\hat{\mathbf{z}}_{j|k}^{\mathrm{T}}+\mathbf{P}_{j|k}\\ \mathbb{E}_{\mathbf{z}_k|\mathbf{Y}_{1:k},\hat{\Theta}_k^{(i)}}\left(\mathbf{z}_j\mathbf{z}_{j-1}^{\mathrm{T}}\right) &= \hat{\mathbf{z}}_{j|k}\hat{\mathbf{z}}_{j-1|k}^{\mathrm{T}}+\mathbf{M}_{j|k}\end{aligned} \tag{21}$$

Now the calculation of E-Step is completed. Then, we start to compute M-Step by (16). For the convenience of derivation, let

$$\Pi_j = \Phi-\Psi\mathbf{A}_j^{\mathrm{T}}-\mathbf{A}_j\Psi^{\mathrm{T}}+\mathbf{A}_j\Sigma\mathbf{A}_j^{\mathrm{T}} = \begin{bmatrix}\varsigma 11(j) & \varsigma 12(j)\\ \varsigma 21(j) & \varsigma 22(j)\end{bmatrix}.$$

The result is given by the following theorem.

Theorem 1: For the model in this paper, the globally unique optimal solution $\hat{\Theta}_k^{(i+1)}$ to (14) is

$$\begin{aligned}\boldsymbol{\mu}_{0k}^{(i+1)} &= \hat{\mathbf{z}}_{0|k}\\ \mathbf{P}_{0k}^{(i+1)} &= \mathbf{P}_{0|k}\end{aligned}$$

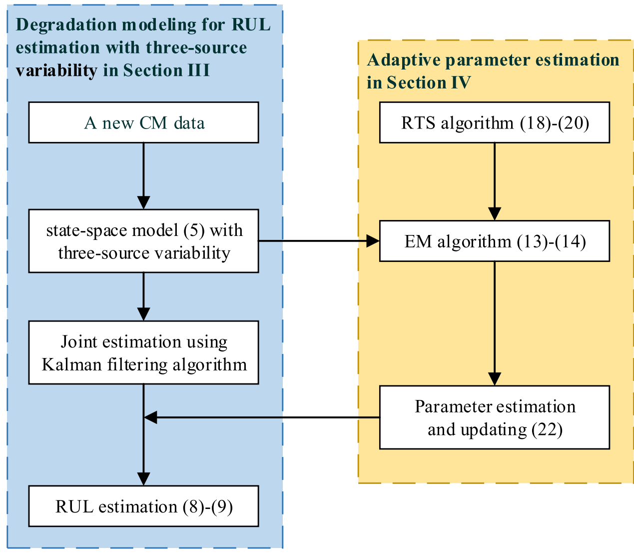


FIGURE 1. Flowchart of the proposed method.

$$\begin{aligned}\left(\sigma^2\right)_k^{(i+1)} &= \frac{1}{k}\sum_{j=1}^{k}\left(\frac{\varsigma 11(j)}{t_j-t_{j-1}}\right)\\ \left(\nu^2\right)_k^{(i+1)} &= \frac{1}{k}\sum_{j=1}^{k}\varsigma 22(j)\\ \left(\gamma^2\right)_k^{(i+1)} &= \frac{1}{k}\sum_{j=1}^{k}\left[\left(y_j-\mathbf{C}\mathbf{z}_{j|k}\right)\left(y_j-\mathbf{C}\mathbf{z}_{j|k}\right)^{\mathrm{T}}+\mathbf{C}\mathbf{P}_{j|k}\mathbf{C}^{\mathrm{T}}\right]\end{aligned} \tag{22}$$

The proof of this theorem is provided in the Appendix.

In sum, the flowchart of the proposed method is shown in Fig. 1.

As shown in Fig. 1, once a new CM data is available, the model parameters can be adaptively estimated and updated in real-time using the RTS algorithm (18)-(20) and the EM algorithm (13)-(14). To estimate the RUL distribution, the Kalman filtering algorithm is used to jointly estimate the underlying degradation state and the drift coefficient. According to the estimated parameters (22) and the above joint estimation results, the RUL is estimated by (8) and (9).

V. EMPIRICAL STUDY

In this section, we provide a practical example of a gyroscope in an inertial navigation system to verify the effectiveness and superiority of the proposed method.

The gyroscope fixed on the inertial platform is a key component of the aerospace and missile weapon system, and its performance directly affects the accuracy of navigation. When the inertial platform is working, the high-speed rotation of the gyroscope's rotor will inevitably cause mechanical wear of the rotating shaft. As the wear accumulates, the drift coefficient increases and eventually leads to the failure. Therefore, accurately estimating the RUL of the gyroscope is critical to improving the safety and reliability of the entire system.

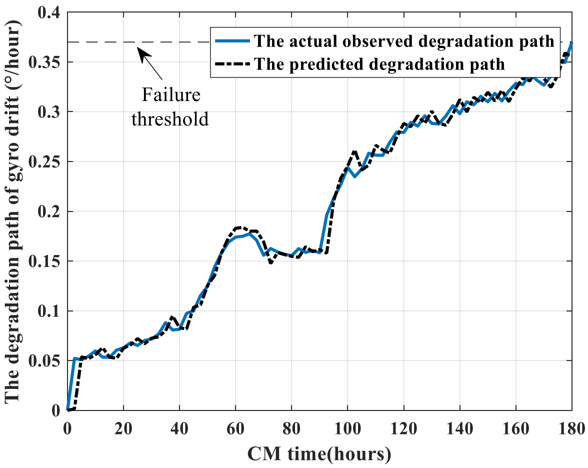


FIGURE 2. **The actual degradation data and the predictions of our model.**

The following is to demonstrate the proposed approach through the CM data by a certain type of the gyroscope in [27]. In the experiment, the fault threshold of the drift coefficient was set as $\omega = 0.37(°/\text{h})$ according to the technical index of this gyroscope, and the health status of the gyroscope was monitored with the sampling interval of 2.5 hours. When the gyroscope was running to 180.5 hours, the drift coefficient reaches the fault threshold for the first time, and its lifetime was considered to be terminated. As shown in Fig. 2, a total of 73 CM points of drift coefficient degradation data were collected during the monitoring period.

Fig. 2 shows that the drift coefficient of the gyroscope increases with the monitoring time as a whole. Using the proposed method, the one-step prediction value of the drift coefficient and the PDF of the estimated RUL at the current time can be obtained at each CM point. To be more specific, the initial parameter vector of the model are set as $\Theta_0 = [0.005, 0.0002, 0.0001, 0.0001, 0.0001, 0.001]^{\mathrm{T}}$, and the one-step prediction of the degradation path by Kalman filter is illustrated in Fig. 2. Obviously, the predicted path is quite close to the actual degraded path, and the root mean square error (RMSE) of the predictions is 4.4281×10^{-5}, which indicates that our developed model can effectively model the gyroscope degradation path. With the accumulation of the CM data, the model parameter vector $\hat{\Theta}_k$, including μ_λ, σ_λ^2, σ_x^2, σ^2, ν^2 and γ^2, are adaptively estimated and updated at each CM point. Correspondingly, the parameter updating process is shown in Fig. 3.

Fig. 3 shows that the model parameters can converge quickly with the accumulation of the CM data. Once the model parameters are updated at each CM point, the PDF and the expectation of the estimated RUL under three-source variability can be calculated by (8) and (9). Furthermore, Fig. 4 shows the results of the proposed method for the RUL estimation, where $\alpha - \lambda$ performance is adopted to quantify the accuracy of the RUL estimation [39], [40].

Fig. 4(a) illustrates the distributions for the RUL estimation at every five CM points and the expectation of the estimated RUL at each CM point. It can be clearly observed from the

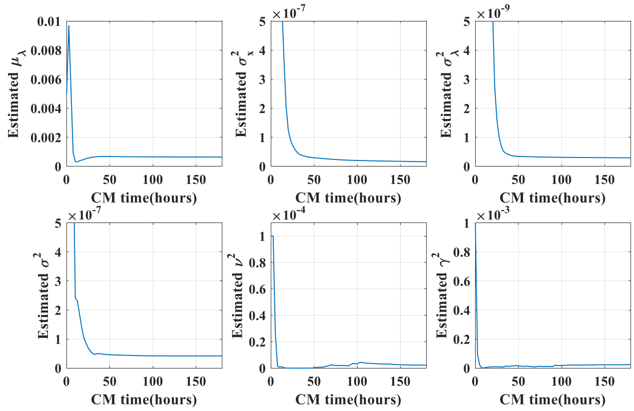


FIGURE 3. **Adaptive updating process of model parameters.**

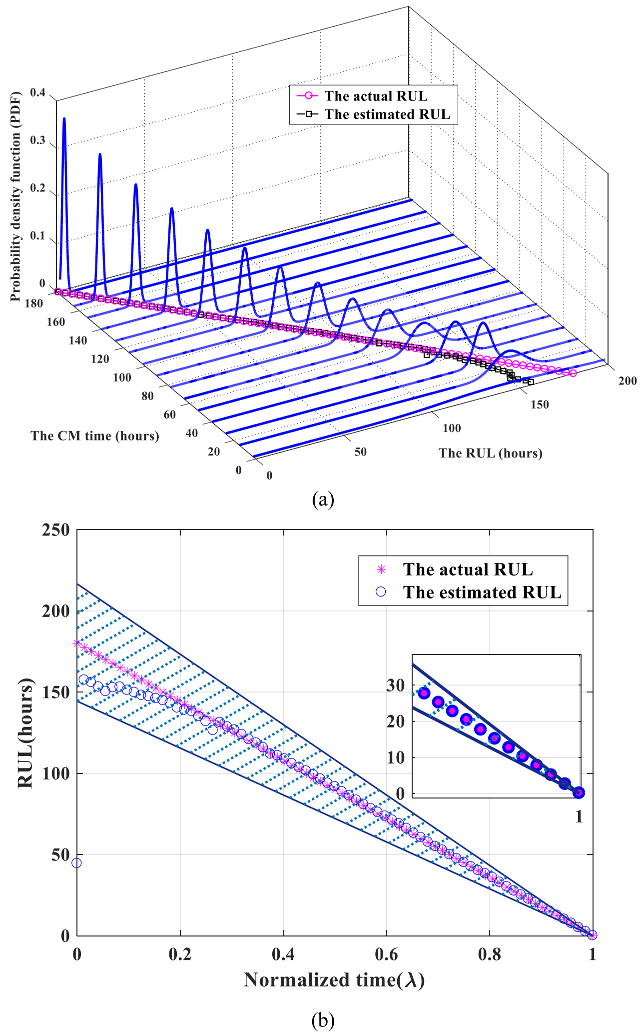


FIGURE 4. **The results of the proposed method. (a) The PDFs of the estimated RUL at different CM points. (b) $\alpha - \lambda$ performance of the RUL estimation ($\alpha = 20\%$).**

Fig. 4 that the estimated RUL is significantly different from the actual RUL due to the lack of degradation data at an early stage of the CM. However, as the CM time increases, the accuracy of the RUL estimation is constantly improved and the PDF curve of the estimated RUL is getting narrower

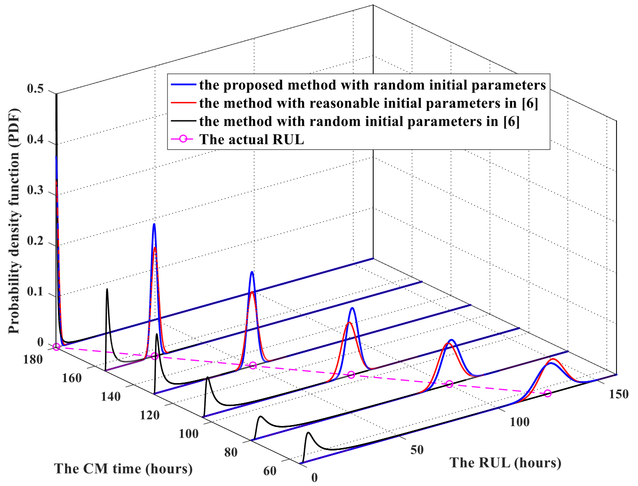


FIGURE 5. Comparative results of the proposed method and the method in [6].

and sharper, which indicates that more and more degradation data are used to estimate the model parameters. Furthermore, the uncertainty of the RUL estimation continues to decrease, which is especially important in the PHM decision-making field [37].

In order to further illustrate the advantages of the proposed method, we compare the proposed method with the method in [6] about their performance in the RUL estimation. It is worth noting that the method in [6] needs initial parameters as well the proposed method. In order to compare the advantages of the two methods, the initial parameters of the proposed method in the following comparison are randomly generated, whereas we consider two cases for the initial parameters of the method in [6]. One is to reasonably select the initial parameters and the other is to randomly generate the initial parameters. Fig. 5 shows the comparison results.

As shown in Fig. 5, for the method of reasonably selecting the initial parameters in [6] and the proposed method of randomly generating the initial parameters, it is intuitive that the variation range of the PDFs of the estimated RULs covers the actual RULs. In addition, at an early stage of the CM, the uncertainty in the estimated RUL of the proposed method is greater than that from the method in [6]. But with the accumulation of the CM data, our estimated PDFs of the RULs are narrower and sharper than the results of the method in [6], which indicates that the estimated RULs of the proposed method are less uncertainty and higher accuracy, especially after the convergence of parameter estimation. However, if the initial parameters of the method in [6] are randomly generated, the RUL estimation results may be incorrect as shown in Fig. 5, i.e., the actual RULs are outside the ranges of PDFs of the estimated RULs.

The reason for the above problems is that the method in [6] is limited by the amount of data, and the initial parameters of the model cannot be estimated in real-time, so that the RUL estimation results depend largely on the selection of the initial parameters. In addition, the method in [6] does not utilize the real-time degradation data of the current working system to update the model parameters. The obtained RUL estimation results are more suitable to describe the common properties of similar systems, and it is difficult to reflect the specific characteristics of the current working system. In comparison, the proposed method has better robustness for the selection of initial parameters of the model, and even the randomly generated initial parameters can get better results. As a result, for the newly developed small-sample systems, since there are no enough historical degradation data to select the initial parameters at the beginning, the proposed method may be more suitable for the RUL estimation.

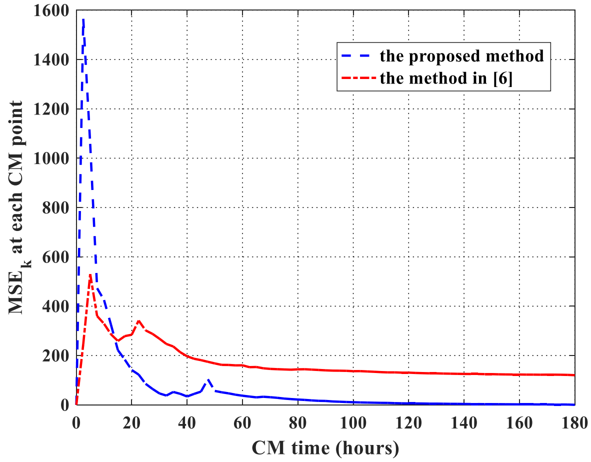


FIGURE 6. The MSE of the estimated RUL at all CM points.

In order to further quantify the comparison results, the mean square error (MSE) is introduced, which takes into account the accuracy of the RUL estimation and the uncertainty of the RUL distribution, and thus it is a commonly used indicator in the RUL prediction, defined as

$$\mathrm{MSE}_k = \int_0^{\infty} (l_k - \mathbb{L}_k)^2 f_{L_k|\mathbf{Y}_{1:k}}(l_k|\mathbf{Y}_{1:k})\mathrm{d}l_k \qquad (23)$$

where $\mathbb{L}_k$ is the actual RUL at t_k time and $f_{L_k|\mathbf{Y}_{1:k}}(l_k|\mathbf{Y}_{1:k})$ is the PDF of the estimated RUL.

Since the method with random initial parameters in [6] will get unsatisfactory results, it is no practical significance to calculate its MSE. Therefore, Fig. 6 shows the MSE comparison of the proposed method and the method in [6] (reasonable selection of model initial parameters) at all CM points. In the early stage of estimation, the calculated MSE results are not satisfactory because the parameter estimate of the proposed method has not yet converged. But with the convergence of parameters and the accumulation of the CM data, the MSE of the proposed method is significantly lower than that from the method in [6]. These comparisons show that the proposed method can effectively overcome the problem of inaccurate parameter estimation caused by the small amount of data, and improve the prognosis accuracy.

VI. CONCLUSION

In this paper, we consider a general linear degradation model based on the Wiener process to simultaneously characterize the temporal variability, unit-to-unit variability and measurement variability in the RUL estimation. In this case, we formulate the PDF and expectation of the estimated RUL. The main contribution of this paper is to present an adaptive parameter estimation method based on the EM algorithm to solve the problem of the lack of historical degradation data in the newly developed small-sample systems. By constructing the state-space model and the proposed parameter estimation method, the RUL estimation results and the model parameters can be adaptively estimated and updated in real-time when a new measurement data is available. Finally, the effectiveness of the proposed method is verified via a practical example of gyroscope in the inertial navigation system. Comparisons with the existing approach show that the proposed approach can effectively overcome the influence of inaccurate parameter estimation caused by less data. In addition, the proposed method can improve the accuracy of the RUL estimation and has good robustness for selecting initial parameters of the model.

There are several directions to be further studied. Firstly, the limitation of the proposed method is that we only consider a general linear degradation model based on the Wiener process, but for complex engineering systems in practice, a nonlinear degradation model may be more appropriate. Secondly, the proposed model is constructed under the assumption of the gradual degradation process, but the failure of complex systems is due to the interaction between internal degradation and external shocks, frequently. Therefore, the proposed model can be extended to a degradation model which incorporates the impact of shocks in the further work. In addition, we primarily focus on the issues associated with the RUL estimation and parameter estimation, but the fundamental purpose of the RUL estimation is to formulate a more reasonable maintenance strategy in practice. Therefore, how to use the results of the RUL estimation to make maintenance decisions scientifically is worthy of further study.

APPENDIX

PROOF OF THEOREM 1

The core goal of the M-step in (14) is to find the parameter vector Θ maximizing $\ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)$. The stationary point can be obtained by $\partial\ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)/\partial\Theta = 0$. In addition, we need to prove that $\hat{\Theta}_k^{(i+1)}$ is the only stationary point of $\ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)$.

Note that the third term of the formulation $\ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)$ in (16), which is the only term depending upon$\mathbf{Q}_j$. However, we cannot get the estimated σ^2 and v^2 in the $(i+1)$th step by directly taking the partial derivative with respect to $\mathbf{Q}_j$ because the matrices $\mathbf{A}_j$ and $\mathbf{Q}_j$are varied with the sampling interval. Therefore, the third term is rewritten as follows:

$$\sum_{j=1}^{k}\left\{\log\left|\mathbf{Q}_j\right| + \mathrm{Tr}\left\{\mathbf{Q}_j^{-1}\left[\Phi-\Psi\mathbf{A}_j^{\mathrm{T}}-\mathbf{A}_j\Psi^{\mathrm{T}}+\mathbf{A}_j\Sigma\mathbf{A}_j^{\mathrm{T}}\right]\right\}\right\}$$
$$=\sum_{j=1}^{k}\left\{\log\left|\mathbf{Q}_j\right| + \mathrm{Tr}\left\{\mathbf{Q}_j^{-1}\Pi_j\right\}\right\}$$
$$=\sum_{j=1}^{k}\left\{\begin{array}{l}\log\left[\sigma^2 v^2\left(t_j-t_{j-1}\right)\right]+ \\ \mathrm{Tr}\left\{\left[\begin{array}{cc}\frac{1}{\sigma^2\left(t_j-t_{j-1}\right)} & 0 \\ 0 & \frac{1}{v^2}\end{array}\right]\left[\begin{array}{cc}\varsigma_{11(j)} & \varsigma_{12(j)} \\ \varsigma_{21(j)} & \varsigma_{22(j)}\end{array}\right]\right\}\end{array}\right\}$$
$$=\sum_{j=1}^{k}\left\{\log\left[\sigma^2 v^2\left(t_j-t_{j-1}\right)\right]+\frac{\varsigma_{11(j)}}{\sigma^2\left(t_j-t_{j-1}\right)}+\frac{\varsigma_{22(j)}}{v^2}\right\} \tag{24}$$

Then we can take the partial derivatives with respect to σ^2 and v^2 in (24), respectively. As a result, in the $(i+1)$th step, we have

$$\begin{aligned}\left(\sigma^2\right)_k^{(i+1)} &= \frac{1}{k}\sum_{j=1}^{k}\left(\frac{\varsigma_{11(j)}}{t_j-t_{j-1}}\right)\\ \left(v^2\right)_k^{(i+1)} &= \frac{1}{k}\sum_{j=1}^{k}\varsigma_{22(j)}\end{aligned} \tag{25}$$

Via a similar method, we can take the partial derivatives with respect to $\boldsymbol{\mu}_0$, $\mathbf{P}_0$, and γ^2 in (16), respectively. Then, we obtain (27)-(29), as shown at the top of the next page. As a result, in the $(i+1)$th step, we have

$$\begin{aligned}\boldsymbol{\mu}_{0k}^{(i+1)} &= \hat{\mathbf{z}}_{0|k}\\ \mathbf{P}_{0k}^{(i+1)} &= \mathbf{P}_{0|k}\\ \left(\gamma^2\right)_k^{(i+1)} &= \frac{1}{k}\sum_{j=1}^{k}\left[\left(y_j-\mathbf{C}\mathbf{z}_{j|k}\right)\left(y_j-\mathbf{C}\mathbf{z}_{j|k}\right)^{\mathrm{T}}+\mathbf{C}\mathbf{P}_{j|k}\mathbf{C}^{\mathrm{T}}\right]\end{aligned} \tag{26}$$

This completes the proof that $\hat{\Theta}_k^{(i+1)}$ is the only stationary point of $\ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)$, i.e., (22). We now prove that $\hat{\Theta}_k^{(i+1)}$is the maximum point of $\ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)$, i.e., the Hessian matrix of $\ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)$ is a negative definite matrix at $\Theta = \hat{\Theta}_k^{(i+1)}$, and hence the second partial derivative $\partial^2\ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)\partial\Theta\partial\Theta^{\mathrm{T}}$ is calculated by (30), as shown at the top of the next page, where

$$\phi = \sum_{j=1}^{k}\left(\frac{\varsigma_{11(j)}}{t_j-t_{j-1}}\right), \quad \psi = \sum_{j=1}^{k}\varsigma_{22(j)},$$
$$\vartheta = \sum_{j=1}^{k}\left[\left(y_j-\mathbf{C}\mathbf{z}_{j|k}\right)\left(y_j-\mathbf{C}\mathbf{z}_{j|k}\right)^{\mathrm{T}}+\mathbf{C}\mathbf{P}_{j|k}\mathbf{C}^{\mathrm{T}}\right]$$

For the notation convenience, let $n_1 = \left(-\mathbf{P}_{0|k}^{-1}, 0, 0, 0, 0\right)$, $n_2 = \left(0, -\frac{1}{2}\mathbf{P}_{0|k}^{-1}\otimes\mathbf{P}_{0|k}^{-1}, 0, 0, 0\right), \ldots, n_5 = \left(0, 0, 0, 0, -\frac{1}{2}\frac{k^3}{[\vartheta]^2}\right)$ denote the rows of the Hessian matrix, respectively. Firstly, we derive the first row of the Hessian matrix, i.e., $n_1 = \left(-\mathbf{P}_{0|k}^{-1}, 0, 0, 0, 0\right)$. Obviously, the partial derivative of (27) with respect to the unknown parameters σ^2, v^2 and γ^2 are equal to zero, respectively. According to the standard

$$\frac{\partial \ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)}{\partial \boldsymbol{\mu}_0} = -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\mu}_0}\mathrm{Tr}\left\{\mathbf{P}_0^{-1}\left[\left(\hat{\mathbf{z}}_{0|k}-\boldsymbol{\mu}_0\right)\left(\hat{\mathbf{z}}_{0|k}-\boldsymbol{\mu}_0\right)^{\mathrm{T}}+\mathbf{P}_{0|k}\right]\right\} = -\mathbf{P}_0^{-1}\left(\boldsymbol{\mu}_0-\hat{\mathbf{z}}_{0|k}\right) \tag{27}$$

$$\frac{\partial \ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)}{\partial \mathbf{P}_0} = -\frac{1}{2}\frac{\partial}{\partial \mathbf{P}_0}\left\{\log|\mathbf{P}_0|+\mathrm{Tr}\left\{\mathbf{P}_0^{-1}\left[\left(\hat{\mathbf{z}}_{0|k}-\boldsymbol{\mu}_0\right)\left(\hat{\mathbf{z}}_{0|k}-\boldsymbol{\mu}_0\right)^{\mathrm{T}}+\mathbf{P}_{0|k}\right]\right\}\right\} = -\frac{1}{2}\left(\mathbf{P}_0^{-1}-\mathbf{P}_0^{-1}\mathbf{P}_{0|k}\mathbf{P}_0^{-1}\right) \tag{28}$$

$$\begin{aligned}\frac{\partial \ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)}{\partial \gamma^2} &= -\frac{1}{2}\frac{\partial}{\partial \gamma^2}\left\{k\log\gamma^2+\mathrm{Tr}\left\{\gamma^{-2}\sum_{j=1}^{k}\left[\left(y_j-\mathbf{C}\mathbf{z}_{j|k}\right)\left(y_j-\mathbf{C}\mathbf{z}_{j|k}\right)^{\mathrm{T}}+\mathbf{C}\mathbf{P}_{j|k}\mathbf{C}^{\mathrm{T}}\right]\right\}\right\}\\ &= -\frac{1}{2}\left\{k\gamma^{-2}-\gamma^{-4}\sum_{j=1}^{k}\left[\left(y_j-\mathbf{C}\mathbf{z}_{j|k}\right)\left(y_j-\mathbf{C}\mathbf{z}_{j|k}\right)^{\mathrm{T}}+\mathbf{C}\mathbf{P}_{j|k}\mathbf{C}^{\mathrm{T}}\right]\right\}\end{aligned} \tag{29}$$

$$\left.\frac{\partial^2 \ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)}{\partial\Theta\partial\Theta^{\mathrm{T}}}\right|_{\Theta=\hat{\Theta}_k^{(i+1)}} = \begin{bmatrix} -\mathbf{P}_{0|k}^{-1} & 0 & 0 & 0 & 0\\ 0 & -\frac{1}{2}\mathbf{P}_{0|k}^{-1}\otimes\mathbf{P}_{0|k}^{-1} & 0 & 0 & 0\\ 0 & 0 & -\frac{1}{2}\frac{k^3}{[\phi]^2} & 0 & 0\\ 0 & 0 & 0 & -\frac{1}{2}\frac{k^3}{[\psi]^2} & 0\\ 0 & 0 & 0 & 0 & -\frac{1}{2}\frac{k^3}{[\vartheta]^2}\end{bmatrix} \tag{30}$$

Kronecker product identities in [35], we further have

$$\left.\frac{\partial^2 \ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)}{\partial \boldsymbol{\mu}_0\partial \boldsymbol{\mu}_0^{\mathrm{T}}}\right|_{\Theta=\hat{\Theta}_k^{(i+1)}} = -\left.\frac{\partial}{\partial \boldsymbol{\mu}_0^{\mathrm{T}}}\mathbf{P}_0^{-1}\left(\boldsymbol{\mu}_0-\hat{\mathbf{z}}_{0|k}\right)\right|_{\Theta=\hat{\Theta}_k^{(i+1)}} = -\mathbf{P}_{0|k}^{-1} < 0 \tag{31}$$

and

$$\begin{aligned}\left.\frac{\partial^2 \ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)}{\partial \boldsymbol{\mu}_0\partial\mathrm{vec}\left\{\mathbf{P}_0\right\}^{\mathrm{T}}}\right|_{\Theta=\hat{\Theta}_k^{(i+1)}} &= -\frac{1}{2}\left.\frac{\partial}{\partial\mathrm{vec}\left\{\mathbf{P}_0\right\}^{\mathrm{T}}}\mathrm{vec}\left\{2\mathbf{P}_0^{-1}\left(\boldsymbol{\mu}_0-\hat{\mathbf{z}}_{0|k}\right)\right\}\right|_{\Theta=\hat{\Theta}_k^{(i+1)}}\\ &= \left[\left(\boldsymbol{\mu}_0-\hat{\mathbf{z}}_{0|k}\right)\otimes\mathbf{I}_{2\times2}\right]\left.\left(\mathbf{P}_0^{-1}\otimes\mathbf{P}_0^{-1}\right)\right|_{\Theta=\hat{\Theta}_k^{(i+1)}}\\ &= 0\end{aligned} \tag{32}$$

This completes the proof of the first row of the Hessian matrix. Other rows are computed in a similar method. Hence, we have the following results.

$$\begin{aligned}\left.\frac{\partial^2 \ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)}{\partial\mathrm{vec}\left\{\mathbf{P}_0\right\}\partial\mathrm{vec}\left\{\mathbf{P}_0\right\}^{\mathrm{T}}}\right|_{\Theta=\hat{\Theta}_k^{(i+1)}} &= -\frac{1}{2}\left.\left\{\begin{array}{l}-\mathbf{P}_0^{-1}\otimes\mathbf{P}_0^{-1}+\\ \left[\left(\mathbf{P}_{0|k}\mathbf{P}_0^{-1}\right)^{\mathrm{T}}\otimes\mathbf{I}_{2\times2}\right]\left(\mathbf{P}_0^{-1}\otimes\mathbf{P}_0^{-1}\right)+\\ \left[\mathbf{I}_{2\times2}\otimes\mathbf{P}_0^{-1}\mathbf{P}_{0|k}\right]\left(\mathbf{P}_0^{-1}\otimes\mathbf{P}_0^{-1}\right)\end{array}\right\}\right|_{\Theta=\hat{\Theta}_k^{(i+1)}}\\ &= -\frac{1}{2}\mathbf{P}_{0|k}^{-1}\otimes\mathbf{P}_{0|k}^{-1}\end{aligned} \tag{33}$$

$$\left.\frac{\partial^2 \ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)}{\partial\left(\sigma^2\right)^2}\right|_{\Theta=\hat{\Theta}_k^{(i+1)}} = -\frac{1}{2}\left.\left(-\frac{k}{\sigma^4}+\frac{2\phi}{\sigma^6}\right)\right|_{\Theta=\hat{\Theta}_k^{(i+1)}} = -\frac{1}{2}\frac{k^3}{[\phi]^2} \tag{34}$$

$$\left.\frac{\partial^2 \ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)}{\partial\left(\nu^2\right)^2}\right|_{\Theta=\hat{\Theta}_k^{(i+1)}} = -\frac{1}{2}\left.\left(-\frac{k}{\nu^4}+\frac{2\psi}{\nu^6}\right)\right|_{\Theta=\hat{\Theta}_k^{(i+1)}} = -\frac{1}{2}\frac{k^3}{[\psi]^2} \tag{35}$$

$$\left.\frac{\partial^2 \ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)}{\partial\left(\gamma^2\right)^2}\right|_{\Theta=\hat{\Theta}_k^{(i+1)}} = -\frac{1}{2}\left.\left(-\frac{k}{\gamma^4}+\frac{2\vartheta}{\gamma^6}\right)\right|_{\Theta=\hat{\Theta}_k^{(i+1)}} = -\frac{1}{2}\frac{k^3}{[\vartheta]^2} \tag{36}$$

This completes the proof of each row of the Hessian matrix. Finally, let us now prove that the Hessian matrix is a negative definite matrix at $\Theta = \hat{\Theta}_k^{(i+1)}$. For this purpose, the order principal minor determinant of the Hessian matrix is calculated as follows:

$$\begin{aligned}&\Delta_1 = -\mathbf{P}_{0|k}^{-1} < 0,\quad \Delta_2 = \frac{1}{2}\mathbf{P}_{0|k}^{-1}\cdot\left(\mathbf{P}_{0|k}^{-1}\otimes\mathbf{P}_{0|k}^{-1}\right) > 0,\\ &\Delta_3 = -\frac{1}{2}\frac{k^3}{[\phi]^2}\Delta_2 < 0,\quad \Delta_4 = -\frac{1}{2}\frac{k^3}{[\psi]^2}\Delta_3 > 0,\\ &\Delta_5 = -\frac{1}{2}\frac{k^3}{[\vartheta]^2}\Delta_4 < 0.\end{aligned} \tag{37}$$

The results indicate that the Hessian matrix is a negative definite matrix at $\Theta = \hat{\Theta}_k^{(i+1)}$ and $\hat{\Theta}_k^{(i+1)}$ is the only stationary point of $\ell\left(\Theta|\hat{\Theta}_k^{(i)}\right)$. As a result, $\hat{\Theta}_k^{(i+1)}$ given by (22) is the globally unique optimal solution.

This completes the proof of Theorem 1.

REFERENCES

[1] N. M. Vichare and M. G. Pecht, "Prognostics and health management of electronics," *IEEE Trans. Compon. Packag. Technol.*, vol. 29, no. 1, pp. 222–229, Mar. 2006.

[2] Z.-S. Ye and M. Xie, "Stochastic modelling and analysis of degradation for highly reliable products," *Appl. Stochastic Models Bus. Ind.*, vol. 31, no. 1, pp. 16–32, 2015.

[3] C. C. Chen, G. Vachtsevanos, and M. E. Orchard, "Machine remaining useful life prediction: An integrated adaptive neuro-fuzzy and high-order particle filtering approach," *Mech. Syst. Signal Process.*, vol. 28, pp. 597–607, 2012.

[4] K. Medjaher, D. A. Tobon-Mejia, and N. Zerhouni, "Remaining useful life estimation of critical components with application to bearings," *IEEE Trans. Rel.*, vol. 61, no. 2, pp. 292–302, Jun. 2012.

[5] 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, no. 1, pp. 1–14, 2011.

[6] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Estimating remaining useful life with three-source variability in degradation modeling," *IEEE Trans. Rel.*, vol. 63, no. 1, pp. 167–190, Mar. 2013.

[7] J. Z. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," *Mech. Syst. Signal Process.*, vol. 25, no. 5, pp. 1803–1836, Jul. 2011.

[8] M. Alamaniotis, A. Grelle, and L. H. Tsoukalas, "Regression to fuzziness method for estimation of remaining useful life in power plant components," *Mech. Syst. Signal Process.*, vol. 48, nos. 1–2, pp. 188–198, 2014.

[9] Q. Zhang, P. W.-T. Tse, X. Wan, and G. Xu, "Remaining useful life estimation for mechanical systems based on similarity of phase space trajectory," *Expert Syst. Appl.*, vol. 42, no. 5, pp. 2353–2360, 2015.

[10] O. F. Eker, F. Camci, and I. K. Jennions, "Physics-based prognostic modelling of filter clogging phenomena," *Mech. Syst. Signal Process.*, vol. 75, pp. 395–412, Jun. 2016.

[11] D. An, N. H. Kim, and J. H. Choi, "Practical options for selecting data-driven or physics-based prognostics algorithms with reviews," *Rel. Eng., Syst. Saf.*, vol. 133, pp. 223–236, Jan. 2015.

[12] M. Jouin, R. Gouriveau, D. Hissel, M.-C. Péra, and N. Zerhouni, "Particle filter-based prognostics: Review, discussion and perspectives," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 2–31, May 2016.

[13] 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.

[14] Z. Zhang, X. Si, C. Hu, and Y. Lei, "Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods," *Eur. J. Oper. Res.*, vol. 271, no. 3, pp. 775–796, Dec. 2018.

[15] S. Qi, Y. Hu, X. Xue, and C. Sun, "Remaining useful life estimation for PEMFC based on monitoring data," in *Proc. 32nd Youth Academic Annu. Conf. Chin. Assoc. Automat. (YAC)*, May 2017, pp. 608–612.

[16] Y. Lei, N. Li, and J. Lin, "A new method based on stochastic process models for machine remaining useful life prediction," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 12, pp. 2671–2684, Dec. 2016.

[17] V. R. Joseph and I.-T. Yu, "Reliability improvement experiments with degradation data," *IEEE Trans. Rel.*, vol. 55, no. 1, pp. 149–157, Mar. 2006.

[18] Z.-X. Zhang, X.-S. Si, C.-H. Hu, Q. Zhang, T.-M. Li, and C.-Q. Xu, "Planning repeated degradation testing for products with three-source variability," *IEEE Trans. Rel.*, vol. 65, no. 2, pp. 640–647, Jun. 2016.

[19] J.-F. Zheng, X.-S. Si, C.-H. Hu, Z.-X. Zhang, and W. Jiang, "A nonlinear prognostic model for degrading systems with three-source variability," *IEEE Trans. Rel.*, vol. 65, no. 2, pp. 736–750, Jun. 2016.

[20] D. Wang and K.-L. Tsui, "Brownian motion with adaptive drift for remaining useful life prediction: Revisited," *Mech. Syst. Signal.*, vol. 99, pp. 691–701, Jan. 2018.

[21] W. Wang, M. Carr, W. Xu, and K. Kobbacy, "A model for residual life prediction based on Brownian motion with an adaptive drift," *Microelectron. Rel.*, vol. 51, no. 2, pp. 285–293, 2011.

[22] W.-A. Yan, B.-W. Song, G.-L. Duan, and Y.-M. Shi, "Real-time reliability evaluation of two-phase Wiener degradation process," *Commun. Statist.-Theory Methods*, vol. 46, no. 1, pp. 176–188, 2017.

[23] J.-X. Zhang, C.-H. Hu, X. He, X.-S. Si, Y. Liu, and D.-H. Zhou, "A novel lifetime estimation method for two-phase degrading systems," *IEEE Trans. Rel.*, to be published.

[24] X. S. Si, C. H. Hu, J. Li, and M, Y, Chen, "Degradation data-driven remaining useful life estimation approach under collaboration between Bayesian updating and EM algorithm," *Pattern Recognit. Artif. Intell.*, vol. 26, no. 4, pp. 357–365, 2013.

[25] M.-L. T. Lee and G. A. Whitmore, "Threshold regression for survival analysis: Modeling event times by a stochastic process reaching a boundary," *Stat. Sci.*, vol. 21, no. 4, pp. 501–513, 2006.

[26] X.-S. Si and D. Zhou, "A generalized result for degradation model-based reliability estimation," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 2, pp. 632–637, Apr. 2014. doi: 10.1109/TASE.2013.2260740.

[27] X.-S. Si, W. Wang, C.-H. Hu, M.-Y. Chen, and D.-H. Zhou, "A Wiener-process-based degradation model with a recursive filter algorithm for remaining useful life estimation," *Mech. Syst. Signal Process.*, vol. 35, nos. 1–2, pp. 219–237, 2013.

[28] C. Y. Peng and S. T. Tseng, "MIS-specification analysis of linear degradation models," *IEEE Trans. Rel.*, vol. 58, no. 3, pp. 444–455, Sep. 2009.

[29] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc., B (Methodol.)*, vol. 39, no. 1, pp. 1–38, 1977.

[30] S. Gibson and B. Ninness, "Robust maximum-likelihood estimation of multivariable dynamic systems," *Automatica*, vol. 41, no. 10, pp. 1667–1682, 2005.

[31] S. Gibson, A. Wills, and B. Ninness, "Maximum-likelihood parameter estimation of bilinear systems," *IEEE Trans. Autom. Control.*, vol. 50, no. 10, pp. 1581–1596, Oct. 2005.

[32] V. Kadirkamanathan and S. R. Anderson, "Maximum-likelihood estimation of delta-domain model parameters from noisy output signals," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3765–3770, Aug. 2008.

[33] A. Wills, B. Ninness, and S. Gibson, "Maximum likelihood estimation of state space models from frequency domain data," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 19–33, Jan. 2009.

[34] H. R. Shumway and S. D. Stoffer, *Time Series Analysis and Its Applications*, 3nd ed. New York, NY, USA: Springer, 2011, pp. 326–344.

[35] R. J. Magnus and H. Neudecker, *Matrix Differential Calculus With Applications in Statistics and Econometrics*, 3nd ed. Tilburg, The Netherlands: Wiley, 2007, pp. 34–36.

[36] N. Chen and K. L. Tsui, "Condition monitoring and remaining useful life prediction using degradation signals: Revisited," *IIE Trans.*, vol. 45, no. 9, pp. 939–952, 2013.

[37] M. G. Pecht, "A prognostics and health management roadmap for information and electronics-rich systems," *IEICE ESS Fundam. Rev.*, vol. 3, no. 4, pp. 4–25, 2010.

[38] Z. Wang, C. Hu, W. Wang, Z. Zhou, and X. Si, "A case study of remaining storage life prediction using stochastic filtering with the influence of condition monitoring," *Rel. Eng. Syst. Saf.*, vol. 132, pp. 186–195, Dec. 2014.

[39] A. Saxena *et al.*, "Metrics for evaluating performance of prognostic techniques," in *Proc. Int. Conf. Prognostics Health Manage.*, Oct. 2008, pp. 1–17.

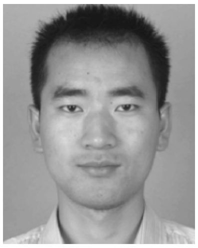
[40] A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel, "Metrics for offline evaluation of prognostic performance," *Int. J. Prognostics Health Manage.*, vol. 1, no. 1, pp. 4–23, 2010.

XI WANG received the B.Eng. degree in control science and management from the Harbin University of Science and Technology, Harbin, China, in 2017. He is currently pursuing the master's degree in instrument science and technology with the Rocket Force University of Engineering, Xi'an, China.

His research interests include prognostics and health management, reliability estimation, predictive maintenance, and lifetime estimation.

CHANGHUA HU received the B.Eng. and M.Eng. degrees from the Rocket Force University of Engineering, Xi'an, China, in 1987 and 1990, respectively, and the Ph.D. degree from the Northwestern Polytechnic University, Xi'an, China, in 1996.

He was a Visiting Scholar with the University of Duisburg, Duisburg, Germany, in 2008. He is currently a Cheung Kong Professor with the Rocket Force University of Engineering. He has authored or coauthored two books and about 100 articles. His research interests include fault diagnosis and prediction, life prognosis, and fault tolerant control.



XIAOSHENG SI received the B.Eng., M.Eng., and Ph.D. degrees from the Department of Automation, Rocket Force University of Engineering, Xi'an, China, in 2006, 2009, and 2014, respectively.

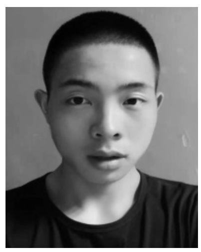
He is currently an Associate Professor with the Rocket Force University of Engineering. He has authored or coauthored more than 50 articles in several journals including the *European Journal of Operational Research*, the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, the IEEE TRANSACTIONS ON RELIABILITY, the IEEE TRANSACTIONS ON FUZZY SYSTEMS, the IEEE TRANSACTIONS ON SYSTEMS, *Man and Cybernetics—Part A*, the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, *Reliability Engineering and System Safety*, and *Mechanical Systems and Signal Processing*. He is an active reviewer for a number of international journals. His research interests include evidence theory, expert systems, prognostics and health management, reliability estimation, predictive maintenance, and lifetime estimation.

Dr. Si is an Associate Editor of the IEEE ACCESS.

JIANFEI ZHENG received the B.Eng., M.Eng., and Ph.D. degrees from the Department of Automation, Rocket Force University of Engineering, Xi'an, China, in 2004, 2008, and 2016, respectively, all in control engineering. He is currently an Assistant Professor with the Rocket Force University of Engineering. He has published over 20 articles in several journals. His research interests include prognostics and health management, predictive maintenance, and inertial navigation systems.

HONG PEI received the B.Eng. degree from the Hefei University of Technology, Hefei, China, in 2014, and the M.Eng. degree from the Rocket Force University of Engineering, Xi'an, China, in 2017. He is currently pursuing the Ph.D. degree. His research interests focus on prognostics and health management, reliability estimation, lifetime estimation, and deep neural networks.

YONG YU received the B.Eng. degree in control science and engineering and the M.Eng. degree in instrument science and technology from the Department of Automation, Rocket Force University of Engineering, Xi'an, China, in 2013, and 2016, respectively. He is currently pursuing the Ph.D. degree. His research interests focus on prognostics and health management, reliability estimation, lifetime estimation, and deep neural networks.



ZIQIANG REN received the B.Eng. degree from Chongqing University, Chongqing, China, in 2017. He is currently pursuing the master's degree in instrument science and technology with the Rocket Force University of Engineering, Xi'an, China.

His research interests focus on prognostics and health management, predictive maintenance, and lifetime estimation.