IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 11, 2020, accepted August 23, 2020, date of publication August 31, 2020, date of current version September 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3020723

Degradation Modeling Based on Wiener Process Considering Multi-Source Heterogeneity

MENG XIAO[1,2], YOUPENG ZHANG[2], YONGJIAN LI[1], AND WENXIAN WANG[1]
[1]School of Railway Transportation, Wuyi University, Jiangmeng 529020, China
[2]School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

Corresponding author: Meng Xiao (westxm@126.com)

ABSTRACT Degradation modeling using heterogeneous degradation data of a system population is critical for prognostics and health management. The heterogeneity may result from the inherent differences in the degrading systems and measuring instruments, as well as from the extrinsic environmental differences, making it difficult to predict the failure. Current studies have not considered the multiple sources of heterogeneity simultaneously, and consequently, have failed to capture the actual degradation process. To explain and quantify the combined effect of multi-source heterogeneity, we present a random effects Wiener process model with heteroscedastic measurement errors, where the parameters of the drift, diffusion, and error variance are assumed as random variables following a certain distribution. The Markov chain Monte Carlo method is adopted to estimate the posterior distributions of the actual degradation states and the model parameters under a hierarchical Bayesian framework. Based on the concept of first hitting time, the failure time distribution is estimated. To verify the presented approach, comparative studies are conducted with simulation and laser degradation data. The results show that considering multi-source heterogeneity can further eliminate the inter-individual variation in degradation data and improve the model fitting and prediction accuracy.

INDEX TERMS Degradation modeling, Wiener process, heterogeneity, Bayesian analysis.

I. INTRODUCTION

Degradation data contain repeated measurements of a degradation process over time in a population of systems. These contain important information about the degradation laws of the population and the degradation characteristics of different individuals. Through fitting the degradation data, a degradation model can be used to extract the degradation laws of such systems, providing a basis for remaining useful life prediction and condition-based maintenance [1]. Therefore, data-driven degradation modeling is critical for prognostics and health management [2]. Because the failure time is predicted by extrapolating the model, the prediction accuracy highly depends on the capability of modeling the degradation data. Because of some unobservable endogenous and exogenous factors [3], the observed degradation paths of systems from the same population typically differ from each other, and exhibit significant heterogeneity. The heterogeneity in the degradation data makes it difficult to model the degradation and leads to uncertainty in the prediction [4]. Current studies have not considered these sources of heterogeneity simultaneously, and consequently, have failed to capture the actual degradation process. In degradation modeling based on stochastic processes, the Wiener process has attracted much attention because of its intuitive physical interpretation and favorable mathematical properties [2], [5]. Moreover, the Wiener process is more suitable for non-monotonically degrading systems than the gamma and inverse Gaussian processes. Therefore, in this study, we focus on degradation modeling considering heterogeneity based on the Wiener process, and analyze the influences of multiple heterogeneity factors on the failure prediction.

Accurate modeling of a degradation process depends critically on our understanding of the degradation heterogeneity. The heterogeneity in the degradation data may originate from multiple sources. One of the heterogeneity factors is the well-known unit-to-unit variability of degradation systems [4]. This variability is mainly due to the inherent differences in such systems, such as the variations in raw materials and manufacturing processes [3]. For modeling the unit-to-unit

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

variability, Peng and Tseng [6] assumed the drift parameter to be unit-specific, and proposed a random-drift Wiener process model for accelerated degradation data. Gebraeel *et al.* [7] established an online updating mechanism for random drift using the Bayesian method to better characterize the individual degradation. Si [8] generalized the random-drift Wiener process model into a nonlinear degradation form to describe the change in the degradation rate with time. When the extrinsic environments in which systems operate are homogeneous, it is reasonable to explain the degradation heterogeneity only in terms of the unit-to-unit variability. Here, the extrinsic environment refers to unobserved exogenous factors that may affect the degradation process, such as the operating condition, usage pattern, and ambient temperature. In practice, however, the extrinsic environments differ from system to system, and the extrinsic environmental differences can also lead to heterogeneity. For example, in accelerated degradation tests, it is found that increasing the environmental stress can increase both the rate and volatility of the degradation [9], [10].

Since the heterogeneity may also arise in response to extrinsic environmental differences, the variability in the drift and diffusion parameters and their correlation should be considered [11]. Wang [12] and Wang *et al.* [13] considered the drift and diffusion parameters as statistically correlated random variables to further describe the degradation heterogeneity using time-homogeneous and multi-time-scale Wiener process models, respectively. Ye *et al.* [11] established a fixed proportional relationship between the random drift parameter and the diffusion parameter to demonstrate that systems with a high degradation rate may have large volatility. Wang *et al.* [14] explained this fixed proportion relationship using the acceleration factor invariant principle, and updated the two parameters with online monitoring information. Under the assumption that the drift and diffusion parameters are unit-specific, the random effects Wiener process model can further explain the degradation heterogeneity due to the variations in the degrading systems and extrinsic environments, and improve the model fitting. However, the above degradation models assume that the actual degradation states of the system can be observed accurately, without considering the influences of measurement errors. Because of imperfect instruments [15], random environments [4], and indirect measurements [16], degradation data are often contaminated by measurement errors. Since degradation data can only partly reflect the actual states of degradation in such situations, the measurement uncertainty should be considered in degradation modeling.

To study the degradation process more accurately with degradation data, the combined effects of the degradation heterogeneity and measurement uncertainty need to be considered simultaneously. Jin *et al.* [10] and Si *et al.* [15] employed a random-drift Wiener process model with measurement errors to predict the remaining useful life of a linear degradation system, respectively. Zheng *et al.* [17] further extended this linear degradation model to nonlinear degradation situations. By introducing the measurement errors into the degradation process model, the robustness of the model used to estimate the actual degradation state could be improved [2], [16]. In such degradation models with measurement errors, the effect of extrinsic environmental differences was not considered, and the parameter of measurement errors was fixed for all systems in the population. However, similar to the individual variability existing in a population of degradation systems, the variability among different measuring instruments may also produce heterogeneous measurement errors, resulting in degradation heterogeneity. The heterogeneous measurement errors are common in practice, particularly when a degradation system has its own measuring instrument. Current studies have not paid enough attention to degradation modeling considering the inherent differences in measuring instruments.

The above analyses show that the heterogeneity in degradation data is inevitable, and may originate from the inherent differences in the degradation systems and measuring instruments, as well as from the differences in the extrinsic environments. However, the existing Wiener-process-based degradation models only considered the heterogeneity of the degradation systems or external environments, ignoring the heteroscedastic measurement errors caused by the differences in the measuring instruments. Therefore, we propose a random effects Wiener process model with heteroscedastic measurement errors for degradation modeling with multi-source heterogeneity. Specifically, the parameters of the drift, diffusion, and error variance are assumed to be unit-specific while interpreting the influence of the multi-source heterogeneity, and are subjected to a certain distribution to quantify the uncertainty due to the heterogeneity. Compared with existing work, this paper mainly contributes to the following aspects:

1) The proposed model can simultaneously perceive and quantify the multi-source heterogeneity from degradation data, thus improving the model fit and the accuracy of the failure time prediction. The model can include many existing models as special cases, and is more general and flexible.

2) A novel Markov chain Monte Carlo (MCMC) algorithm is presented to estimate the posterior distributions of the model parameters under a Bayesian hierarchical framework. This algorithm can effectively deal with the complex parameter structure in the proposed model. Its accurateness and robustness are verified by a simulation experiment.

3) A real case is used to analyze the combined effect of multi-source heterogeneity on the estimates of the actual degradation states and the failure time distribution. The results verify the rationality and effectiveness of the proposed method.

The remainder of this article is organized as follows. Section 2 introduces a degradation model with multi-source heterogeneity. A Bayesian parameter estimation method is presented in Section 3. Section 4 gives the failure time distribution based on the posterior estimates of the model parameters. Section 5 provides a simulation study and case study conducted to demonstrate the effectiveness of our

proposed model and algorithm. Finally, we conclude our findings in Section 6.

II. DEGRADATION MODEL DESCRIPTION

We introduce a model based on the Wiener process to describe the degradation process $\{X(t), t \geq 0\}$. The degradation state at time t can be expressed as:

$$X(t) = X(0) + \lambda t + \eta^{-1/2} B(t), \tag{1}$$

where λ is the drift coefficient reflecting the degradation rate, $\eta^{-1/2}$ is the diffusion coefficient reflecting the degradation volatility, and $B(t)$ represents the standard Brownian motion. Without loss of generality, we assume that the initial value of the degradation process is $X(0) = 0$.

Because of imperfect measurements or disturbances from external environment, random measurement errors are inevitable in degradation data. As a result of the measurement errors, degradation data can only partially reflect the actual degradation states, and there exits measurement uncertainty [15]. Let $\{Y(t), t \geq 0\}$ represent the observation sequence of a degradation process; then, the measurement equation can be written as:

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

where $\varepsilon(t)$ is the Gaussian white noise, i.e., $\varepsilon(t) \sim_{i.i.d.} \mathrm{N}(0, \gamma^{-1})$, and $\mathrm{N}(\cdot)$ represents the Gaussian distribution; γ^{-1} represents the error variance. It is further assumed that $\varepsilon(t)$ and $B(t)$ are mutually and statistically independent. These assumptions have been widely adopted in studies on degradation modeling [6], [10], [16], [17].

From a physical point of view, the deterministic component λt in (1) denotes the global degradation trend, the stochastic component $\eta^{-1/2} B(t)$ describes the random fluctuation around the deterministic trend, and accordingly, the drift λ and diffusion η can capture the rate and volatility of degradation respectively. Moreover, the error γ can further explain the uncertain part introduced by measurements in the random fluctuation. Equations (1) and (2) constitute the basic linear degradation model with measurement errors, considering both the time-varying and measurement uncertainties in the degradation process [2]. This model and its nonlinear extension, i.e., the time-varying degradation rate, have been applied to the degradation modeling of battery data [10], LED data [16], and fatigue-crack data [17].

As described in the introduction, the degradation heterogeneity due to the variations in the degradation systems and extrinsic environments can be described by assuming the parameters λ and η to be unit-specific. This forms the basis of the random effects Wiener process model. However, the variability among different measuring instruments may also produce heterogeneous measurement errors, leading to heterogeneity, as expressed in (2). Therefore, we consider a more general case of degradation heterogeneity, where the drift λ, diffusion η, and error γ differ among individuals, and describe the multi-source heterogeneity by a random effects Wiener process model with heteroscedastic measurement errors. Specifically, let (λ, η, γ) be a random vector with a probability density function (PDF) as $p(\lambda, \eta, \gamma|\boldsymbol{\Psi})$, where $\boldsymbol{\Psi}$ is the density parameter. If there are N systems in a population, we may regard the individual degradation parameters $\{\lambda_i, \eta_i, \gamma_i\}$ of each system as independent identically distributed samples from the population distribution $p(\lambda, \eta, \gamma|\boldsymbol{\Psi})$. It can be expressed as:

$$\{\lambda_i, \eta_i, \gamma_i\} \sim_{i.i.d.} p(\lambda, \eta, \gamma|\boldsymbol{\Psi}), \quad i = 1, 2, \ldots, N.$$

Evidently, the above model, which we refer to as model M0, is a general and flexible form that can include the existing models, which considers only one- or two-source heterogeneity, as special cases.

When the parameter γ is a positive constant for each system across the population, model M0 reduces to the random effects Wiener process model with homoscedastic measurement errors, which we refer to as model M1. In particular, when $\gamma \to +\infty$ (i.e., the error variance $\gamma^{-1} \approx 0$), model M0 reduces to the random effects Wiener process model proposed in [12]–[14], where only the inherent differences in the degradation systems and extrinsic environmental differences were considered.

When both the parameters γ and η are positive constants, model M0 reduces to the random-drift Wiener process model with homoscedastic measurement errors proposed in [10], [15], which we refer to as model M2. The model M2 considers only the inherent differences in the degradation systems.

Let $\mathbf{y}_i = [y_{i,1}, y_{i,2}, \ldots, y_{i,n_i}]^{\mathrm{T}}$ and $\mathbf{t}_i = [t_{i,1}, t_{i,2}, \ldots, t_{i,n_i}]^{\mathrm{T}}$ be the degradation measurements and the corresponding measurement time of the ith system, respectively, where $y_{i,j} = Y(t_{i,j})$ denotes the degradation data at the time $t_{i,j}$, and n_i is the number of measurements. Based on the above assumptions of model M0, the problem of model identification is to estimate the parameter $\boldsymbol{\Psi}$ with the degradation measurements $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\}$.

Although considering a greater number of model parameters as random variables can enhance the ability of model description, it increases the complexity of statistical inference. Therefore, an effective statistical inference algorithm is required to estimate the model parameter $\boldsymbol{\Psi}$.

III. PARAMETER ESTIMATION

The parameters of a probabilistic model can be obtained using the MLE or Bayesian method, and the specific selection factors include the model complexity, sample size, and subsequent statistical reasoning based on the estimated parameters. Compared with the MLE, the Bayesian method can directly provide the posterior distribution of the parameter, which is conducive to analyzing the uncertainty in estimating and updating the parameters with online data [7]. In addition, given the complexity of the model and the situation of small-scale samples, the Bayesian method can additionally avoid the high-dimensional optimization and the positive semidefinite problem of the covariance matrix observed in the MLE.

In this study, the Bayesian method is used for the posterior estimation of the parameter $\boldsymbol{\Psi}$.

For ease of presentation, let $\boldsymbol{\theta}_i = \{\lambda_i, \eta_i, \gamma_i\}$ denote the individual parameter of the ith system. The degradation state of the ith system is defined as $\mathbf{x}_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,n_i}]^{\mathrm{T}}$, where $x_{i,j} = X(t_{i,j})$ denotes the degradation state at the time $t_{i,j}$. Based on the conditional normal properties of (1) and (2), for the system $i = 1, 2, \ldots, N$, the degradation state increments $\Delta x_{i,j} = x_{i,j} - x_{i,j-1}$ and the degradation measurements follow normal distributions under the given model parameters $\boldsymbol{\theta}_i$, and can be expressed as:

$$\begin{aligned} \Delta x_{i,j}|\boldsymbol{\theta}_i &\sim \mathrm{N}(\lambda_i \Delta t_{i,j}, \eta_i^{-1} \Delta t_{i,j}), \\ y_{i,j}|x_{i,j}, \boldsymbol{\theta}_i &\sim \mathrm{N}(x_{i,j}, \gamma_i^{-1}), \end{aligned} \tag{3}$$

where $\Delta t_{i,j} = t_{i,j} - t_{i,j-1}$ denotes the interval time between the adjacent observations, $j = 1, 2, \ldots, n_i$.

Based on the statistically independent property of the increments in the Wiener process and the conditional independent assumption of the degradation data, the likelihood function of $\boldsymbol{\theta}_i$ can be expressed as:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}_i) &= p(\mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\theta}_i) p(\mathbf{x}_i|\boldsymbol{\theta}_i) \\ &\propto \prod_{j=1}^{n_i} \eta_i^{1/2} \exp\left(\frac{\eta_i(\Delta x_{i,j} - \lambda_i \Delta t_{i,j})^2}{-2\Delta t_{i,j}}\right) \\ &\quad \times \gamma_i^{1/2} \exp\left(\frac{\gamma_i (y_{i,j} - x_{i,j})^2}{-2}\right). \end{aligned} \tag{4}$$

Since the degradation data $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\}$ of N systems are independent samples, the posterior distribution of the model parameter $\boldsymbol{\Psi}$ is as follows:

$$\begin{aligned} p(\boldsymbol{\Psi}|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N) &\propto p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N|\boldsymbol{\Psi}) p(\boldsymbol{\Psi}) \\ &= \left\{ \prod_{i=1}^{N} \iint \mathcal{L}(\boldsymbol{\theta}_i) p(\boldsymbol{\theta}_i|\boldsymbol{\Psi}) d\mathbf{x}_i d\boldsymbol{\theta}_i \right\} p(\boldsymbol{\Psi}). \end{aligned} \tag{5}$$

Because (5) contains a high-dimensional integral of the state vector $\mathbf{x}_i$ and the individual parameter vector $\boldsymbol{\theta}_i$, it is generally difficult to obtain the posterior distribution of the parameter $\boldsymbol{\Psi}$ analytically except in the case of linear Gaussian distributions. Therefore, the MCMC method is often used in Bayesian analyses to randomly simulate the posterior distribution of unknown parameters, and statistical inference is then implemented based on the results of the random simulation. A Bayesian analysis mainly involves two steps: specifying the prior distributions for the parameters of interest and reasoning its posterior distributions. In the following part, we specify the prior and propose an MCMC algorithm for the model parameter $\boldsymbol{\Psi}$.

A. HIERARCHICAL PRIOR STRUCTURE

It is important to specify an appropriate prior distribution in Bayesian analyses. To facilitate the derivation of the posterior distribution, a conjugate prior distribution is often used. First, we consider the prior distribution $p(\boldsymbol{\theta}|\boldsymbol{\Psi})$ of an individual parameter $\boldsymbol{\theta}$. It is assumed that the measurement error process $\{\varepsilon(t), t \geq 0\}$ is independent of the degradation state process $\{X(t), t \geq 0\}$, i.e., the system degradation behavior does not interact with the degradation measurement. This assumption is also consistent with the non-damage or indirect measurements often taken in engineering practice, such as the vibration signals of bearings and the current or voltage induction signals of batteries [10]. Therefore, the parameters $\{\lambda, \eta\}$ of the degradation process are independent of the parameter γ of the measurement. Moreover, because the degradation rate and degradation volatility can be affected by some of the same factors such as manufacturing materials or extrinsic environments [9], [13], a statistical correlation may be considered between the drift λ and the diffusion η. Based on the above correlation analysis of the individual parameter, the prior is given as:

$$p(\boldsymbol{\theta}|\boldsymbol{\Psi}) = p(\lambda, \eta|\boldsymbol{\Psi}) p(\gamma|\boldsymbol{\Psi}) = p(\lambda|\eta, \boldsymbol{\Psi}) p(\eta|\boldsymbol{\Psi}) p(\gamma|\boldsymbol{\Psi}).$$

Observing the likelihood function (4), we find that λ and η are present in the mean and variance terms of the normal distributions of the degradation increments, respectively, and γ is present in the variance term of the normal distributions of the degradation measurements. In a Bayesian analysis, the normal-gamma distribution is often used as the conjugate prior of the mean and precision parameters of the normal distribution, and the gamma distribution is used as the conjugate prior of the precision parameter of the normal distribution [18]. Therefore, the conjugate prior of the individual parameter $\boldsymbol{\theta}$ can be expressed as [19]

$$\begin{cases} \lambda|\eta \sim \mathrm{N}(\mu, (\kappa\eta)^{-1}), \\ \eta \sim \mathrm{G}(v_\eta, v_\eta / u_\eta), \\ \gamma \sim \mathrm{G}(v_\gamma, v_\gamma / u_\gamma), \end{cases} \tag{6}$$

where $\mathrm{G}(\cdot)$ represents the gamma distribution, which has the PDF:

$$p(x|\alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} \exp(-\beta x), \quad x > 0, \tag{7}$$

with mean α/β, and variance α/β^2.

At this point, the parameter to be estimated is $\boldsymbol{\Psi} = \{\mu, \kappa, v_\eta, u_\eta, v_\gamma, u_\gamma\}$. Further, we can specify the prior of the model parameter $\boldsymbol{\Psi}$ as [19]:

$$\begin{cases} \mu \sim \mathrm{N}(\omega, q), & \kappa \sim \mathrm{G}(a, b), \\ v_\eta \sim \mathrm{E}(\xi_\eta), & u_\eta \sim \mathrm{IG}(\alpha_\eta, \beta_\eta), \\ v_\gamma \sim \mathrm{E}(\xi_\gamma), & u_\gamma \sim \mathrm{IG}(\alpha_\gamma, \beta_\gamma), \end{cases} \tag{8}$$

where $\mathrm{E}(\cdot)$ represents the exponential distribution; $\mathrm{IG}(\cdot)$ represents the inverse gamma distribution with the PDF:

$$p(x|\alpha, \beta) = \frac{\beta^\alpha x^{-\alpha-1}}{\Gamma(\alpha)} \exp\left(-\frac{\beta}{x}\right), \quad x > 0.$$

In the Bayesian context, the parameters $\{\omega, q, a, b, \xi_\eta, \alpha_\eta, \beta_\eta, \xi_\gamma, \alpha_\gamma, \beta_\gamma\}$ are hyper-parameters of the prior of $\boldsymbol{\Psi}$, which are usually set based on prior knowledge or in a diffuse prior form. To sum up, the hierarchical Bayesian degradation

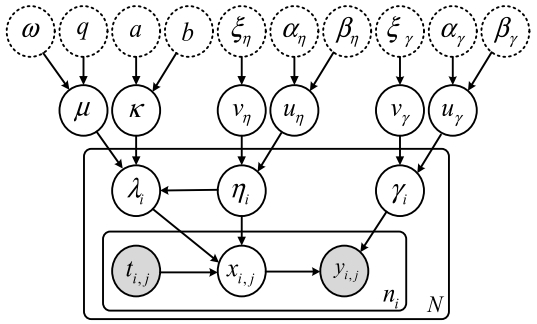

FIGURE 1. Probability structure diagram of a degradation model.

model composed of (1), (2), (6), and (8) can be summarized in the form of a probability structure diagram as shown in Fig. 1. As shown, the shadow node represents an observable variable; the dashed circle node represents a hyperparameter; the other nodes represent unknown parameters or variables; the directed edge represents statistical dependency between the nodes, and the rounded rectangular box represents the number of independent observations of a parameter or variable.

Corresponding to the conditional independent relationship shown in Fig. 1, according to the Bayesian theorem, the joint posterior distribution of all the unknown parameters $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N, \boldsymbol{\Psi}\}$, and the variables $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ can be obtained via Bayes' theorem as

$$\begin{aligned}
&p(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N, \boldsymbol{\Psi}|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N) \\
&\propto \left\{\prod_{i=1}^{N} \mathcal{L}(\boldsymbol{\theta}_i) p(\boldsymbol{\theta}_i|\boldsymbol{\Psi})\right\} p(\boldsymbol{\Psi}) \\
&\propto \left\{\prod_{i=1}^{N}\left[\prod_{j=1}^{n_i} \eta_i^{1/2} \exp\left(\frac{\eta_i(\Delta x_{i,j} - \lambda_i \Delta t_{i,j})^2}{-2\Delta t_{i,j}}\right)\right.\right. \\
&\quad \left.\left. \times\ \gamma_i^{1/2} \exp\left(\frac{\gamma_i(y_{i,j} - x_{i,j})^2}{-2}\right)\right] (\kappa\eta_i)^{1/2} \exp\left(\frac{\kappa\eta_i(\lambda_i - \mu)^2}{-2}\right)\right. \\
&\quad \times \frac{(v_\eta/u_\eta)^{v_\eta} \eta_i^{v_\eta - 1}}{\Gamma(v_\eta)} \exp\left(-\frac{v_\eta}{u_\eta}\eta_i\right) \\
&\quad \left.\times \frac{(v_\gamma/u_\gamma)^{v_\gamma} \gamma_i^{v_\gamma - 1}}{\Gamma(v_\gamma)} \exp\left(-\frac{v_\gamma}{u_\gamma}\gamma_i\right)\right\} q^{1/2} \exp\left(\frac{q(\mu - \omega)^2}{-2}\right) \\
&\quad \times \kappa^{a-1} \exp(-b\kappa)\ \times \exp(-\xi_\eta v_\eta) u_\eta^{-\alpha_\eta - 1} \exp\left(-\frac{\beta_\eta}{u_\eta}\right) \\
&\quad \times \exp(-\xi_\gamma v_\gamma) u_\gamma^{-\alpha_\gamma - 1} \exp\left(-\frac{\beta_\gamma}{u_\gamma}\right). \qquad (9)
\end{aligned}$$

Because of the hierarchical prior structure shown in Fig. 1, the joint posterior represented by (9) is too complex to easily obtain the marginal posterior of the parameter $\boldsymbol{\Psi}$ using the integral method. Therefore, we resort to the MCMC method to randomly simulate its marginal posterior distribution.

B. POSTERIOR INFERENCE

In the MCMC method, an ergodic Markov chain $\{\mathbf{x}^{(k)}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\Psi}^{(k)}\}$, $k = 1,2,\ldots$, is constructed with the joint posterior distribution $p(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\Psi}|\mathbf{y})$ as the stationary distribution, and random samples of the posterior distribution are simulated. Because the MCMC method has good properties, such as consistency and asymptotic normality, it is a reasonable approximation to the posterior distribution [20].

In the hierarchical prior structure expressed in (6) and (8), although the conjugate priors are used as much as possible, the conditional posterior distributions of some parameters are still unknown, and random sampling cannot be carried out directly. Therefore, we use a Metropolis-within-Gibbs algorithm to randomly sample the conditional posterior distributions. This algorithm combines the generality of the Metropolis–Hastings algorithm with the efficiency and convenience of the Gibbs algorithm. For a parameter that cannot be directly sampled from its conditional posterior distribution, the sample may be updated in the Metropolis step. Based on the joint posterior distribution given by (9), the conditional posterior distributions of each degradation state and unknown parameter are derived as follows.

1) CONDITIONAL POSTERIOR DISTRIBUTION OF $x_{i,j}$

Let $\mathbf{x}_{i,(j)} = \{x_{i,1}, \ldots, x_{i,j-1}, x_{i,j+1}, \ldots, x_{i,n_i}\}$ denote the state sequence after removing $x_{i,j}$. The conditional posterior distribution of $x_{i,j}$, for $i = 1, 2, \ldots, N, j = 1, 2, \ldots, (n_i - 1)$, is given by

$$\begin{aligned}
&p(x_{i,j}|\mathbf{x}_{i,(j)}, \mathbf{y}_i, \boldsymbol{\theta}_i) \\
&\propto \exp\left(\frac{\eta_i(\Delta x_{i,j} - \lambda_i \Delta t_{i,j})^2}{-2\Delta t_{i,j}}\right) \\
&\quad \times \exp\left(\frac{\eta_i(\Delta x_{i,j+1} - \lambda_i \Delta t_{i,j+1})^2}{-2\Delta t_{i,j+1}}\right) \exp\left(\frac{\gamma_i(y_{i,j} - x_{i,j})^2}{-2}\right) \\
&\sim \mathrm{N}(x|d_{x_{i,j}}, D_{x_{i,j}}). \qquad (10)
\end{aligned}$$

According to (10), this conditional posterior is still normal, and its conditional variance and mean are respectively:

$$\begin{aligned}
D_{x_{i,j}}^{-1} &= \eta_i(\Delta t_{i,j}^{-1} + \Delta t_{i,j+1}^{-1}) + \gamma_i, \\
d_{x_{i,j}} &= D_{x_{i,j}} \left(\eta_i(x_{i,j}\Delta t_{i,j-1}^{-1} + x_{i,j+1}\Delta t_{i,j+1}^{-1}) + \gamma_i y_{i,j}\right).
\end{aligned}$$

For the last state $x_{i,n_i}, i = 1, 2, \ldots, N$, its conditional posterior distribution is as follows:

$$\begin{aligned}
&p(x_{i,n_i}|\mathbf{x}_{i,(n_i)}, \mathbf{y}_i, \boldsymbol{\theta}_i) \\
&\propto \exp\left(\frac{\eta_i(x_{i,n_i} - x_{i,n_i-1} - \lambda_i \Delta t_{i,n_i})^2}{-2\Delta t_{i,n_i}}\right) \times \exp\left(\frac{\gamma_i(y_{i,n_i} - x_{i,n_i})^2}{-2}\right) \\
&\sim \mathrm{N}(x_{i,n_i}|d_{x_{i,n_i}}, D_{x_{i,n_i}}), \qquad (11)
\end{aligned}$$

where

$$\begin{aligned}
D_{x_{i,n_i}}^{-1} &= \eta_i \Delta t_{i,n_i}^{-1} + \gamma_i, \\
d_{x_{i,n_i}} &= D_{x_{i,1}} \left(\eta_i(x_{i,n_i-1}\Delta t_{i,n_i}^{-1} + \lambda_i) + \gamma_i y_{i,n_i}\right).
\end{aligned}$$

2) CONDITIONAL POSTERIOR DISTRIBUTION OF λ_i

The conditional posterior distribution of λ_i, for $i = 1, 2, \ldots, N$, is

$$\begin{aligned} &p(\lambda_i|\mathbf{x}_i, \mathbf{y}_i, \eta_i, \gamma_i, \boldsymbol{\Psi}) \\ &\propto \prod_{j=1}^{n_i} \exp\left(\frac{\eta_i(\Delta x_{i,j} - \lambda_i \Delta t_{i,j})^2}{-2\Delta t_{i,j}}\right) \exp\left(\frac{\kappa\eta_i(\lambda_i-\mu)^2}{-2}\right) \\ &\sim \mathrm{N}(\lambda_i|d_{\lambda_i}, D_{\lambda_i}), \end{aligned} \tag{12}$$

where $D_{\lambda_i}^{-1} = \eta_i(t_{i,n_i} + \kappa)$, $d_{\lambda_i} = D_{\lambda_i}\eta_i(x_{i,n_i} + \kappa\mu)$.

3) CONDITIONAL POSTERIOR DISTRIBUTION OF η_i

The conditional posterior distribution of η_i, $i = 1, 2, \ldots, N$, is obtained as follows

$$\begin{aligned} &p(\eta_i|\,\mathbf{x}_i, \mathbf{y}_i, \lambda_i, \gamma_i, \boldsymbol{\Psi}) \\ &\propto \prod_{j=1}^{n_i} \eta_i^{1/2} \exp\left(\frac{\eta_i(\Delta x_{i,j} - \lambda_i \Delta t_{i,j})^2}{-2\Delta t_{i,j}}\right)(\kappa\eta_i)^{1/2} \\ &\quad \times \exp\left(\frac{\kappa\eta_i(\lambda_i-\mu)^2}{-2}\right) \eta_i^{v_\eta - 1} \exp\left(-\frac{v_\eta}{u_\eta}\eta_i\right) \\ &\sim \mathrm{G}(\eta_i|d_{\eta_i}, D_{\eta_i}), \end{aligned} \tag{13}$$

where

$$\begin{aligned} D_{\eta_i} &= \sum_{j=1}^{n_i} \frac{(\Delta x_{i,j} - \lambda_i \Delta t_{i,j})^2}{2\Delta t_{i,j}} + \frac{\kappa(\lambda_i-\mu)^2}{2} + \frac{v_\eta}{u_\eta}, \\ d_{\eta_i} &= v_\eta + (n_i+1)/2. \end{aligned}$$

4) CONDITION POSTERIOR DISTRIBUTION OF γ_i

The conditional posterior distribution of γ_i, for $i = 1, 2, \ldots, N$, has the following form:

$$\begin{aligned} &p(\gamma_i|\mathbf{x}_i, \mathbf{y}_i, \lambda_i, \eta_i, \boldsymbol{\Psi}) \\ &\propto \prod_{j=1}^{n_i} \gamma_i^{1/2} \exp\left(\frac{\gamma_i(y_{i,j} - x_{i,j})^2}{-2}\right) \gamma_i^{v_\gamma - 1} \exp\left(-\frac{v_\gamma}{u_\gamma}\gamma_i\right) \\ &\sim \mathrm{G}(\gamma|d_{\gamma_i}, D_{\gamma_i}), \end{aligned} \tag{14}$$

where

$$D_{\gamma_i} = \sum_{j=1}^{n_i} \frac{(y_{i,j} - x_{i,j})^2}{2} + \frac{v_\gamma}{u_\gamma}, \quad d_{\gamma_i} = v_\gamma + n_i/2.$$

5) CONDITIONAL POSTERIOR DISTRIBUTION OF HYPER-PARAMETERS μ AND κ

The conditional posterior distribution of the parameter μ is as follows

$$\begin{aligned} &p(\mu|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N, \kappa) \\ &\propto \prod_{i=1}^{N} \exp\left(\frac{\kappa\eta_i(\lambda_i-\mu)^2}{-2}\right) \exp\left(\frac{q(\mu-\omega)^2}{-2}\right) \\ &\sim \mathrm{N}(\mu|d_\mu, D_\mu), \end{aligned} \tag{15}$$

where

$$D_\mu^{-1} = \kappa \sum_{i=1}^{N} \eta_i + q, \quad d_\mu = D_\mu(\kappa \sum_{i=1}^{N} \eta_i\lambda_i + q\omega).$$

For parameter κ, its formulation is as follows:

$$\begin{aligned} &p(\kappa\,|\,\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N, \mu) \\ &\propto \prod_{i=1}^{N} (\kappa\eta_i)^{1/2} \exp\left(\frac{\kappa\eta_i(\lambda_i-\mu)^2}{-2}\right) \kappa^{a-1} \exp(-b\kappa) \\ &\sim \mathrm{G}(\kappa\,|\,d_\kappa, D_\kappa), \end{aligned} \tag{16}$$

where

$$d_\kappa = a + N/2, \quad D_\kappa = \sum_{i=1}^{N} \frac{\eta_i(\lambda_i-\mu)^2}{2} + b.$$

6) CONDITIONAL POSTERIOR DISTRIBUTIONS OF HYPER-PARAMETERS u_η AND v_η

The conditional posterior distribution of u_η is as follows

$$\begin{aligned} &p(u_\eta\big|\,\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N, v_\eta) \\ &\propto \prod_{i=1}^{N} (u_\eta)^{-v_\eta} \exp\left(\frac{-v_\eta}{u_\eta}\eta_i\right) u_\eta^{-\alpha_\eta - 1} \exp\left(\frac{-\beta_\eta}{u_\eta}\right) \\ &\sim \mathrm{IG}(u_\eta\big|\,d_{u_\eta}, D_{u_\eta}), \end{aligned} \tag{17}$$

where $D_{u_\eta} = v_\eta \sum_{i=1}^{N} \eta_i + \beta_\eta$, $d_{u_\eta} = Nv_\eta + \alpha_\eta$.

The v_η can be calculated using the formulation:

$$\begin{aligned} &p(v_\eta|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N, u_\eta) \\ &\propto \prod_{i=1}^{N} \frac{(v_\eta/u_\eta)^{v_\eta} \eta_i^{v_\eta - 1}}{\Gamma(v_\eta)} \exp\left(-\frac{v_\eta}{u_\eta}\eta_i\right) \exp(-\xi_\eta v_\eta) \\ &= \frac{(v_\eta/u_\eta)^{Nv_\eta}}{\Gamma(v_\eta)^N} \exp\left\{-\left[\sum_{i=1}^{N}\left(\frac{\eta_i}{u_\eta} - \log\eta_i\right) + \xi_\eta\right] v_\eta\right\}, \end{aligned} \tag{18}$$

The distribution of v_η given by (18) is an unknown distribution; thus, it cannot be sampled directly. At this point, a random sampling of this unknown distribution is obtained via the Metropolis step. When the value of v_η in (18) is high, its PDF is similar to a gamma distribution, so the gamma distribution may be chosen as the proposal distribution to generate a candidate [19]. Let $\mathrm{G}(r, r/v_\eta)$ be the proposal gamma distribution; then, the acceptance probability of the candidate $\tilde{v}_\eta$, i.e., $\tilde{v}_\eta \sim \mathrm{G}(r, r/v_\eta)$, can be calculated as

$$\min\left\{1, \frac{p(\tilde{v}_\eta|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N, u_\eta) \times \mathrm{G}(\tilde{v}_\eta|r, r/v'_\eta)}{p(v'_\eta|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N, u_\eta) \times \mathrm{G}(v'_\eta|r, r/\tilde{v}_\eta)}\right\},$$

where v'_η is the sampling result of the previous iteration; and r is the parameter of the proposal distribution. If the candidate $\tilde{v}_\eta$ is rejected, v'_η is still retained for the next iteration. Evidently, v'_η and v'^2_η/r are the mean and variance of the proposal distribution, respectively. The acceptance probability of a candidate can be improved by adjusting the value of r. We selected $r = 10$ for the subsequent experimental research.

7) CONDITIONAL POSTERIOR DISTRIBUTIONS OF HYPER-PARAMETERS u_γ AND v_γ

Similar to u_η and v_η, the conditional posterior distributions of u_γ and v_γ are respectively:

$$\begin{aligned} &p(u_\gamma|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N, v_\gamma) \\ &\propto \prod_{i=1}^{N} (u_\gamma)^{-v_\gamma} \exp\left(\frac{v_\gamma}{-u_\gamma}\gamma_i\right) u_\gamma^{-\alpha_\gamma-1} \exp\left(\frac{\beta_\gamma}{-u_\gamma}\right) \\ &\sim \mathrm{IG}(u_\gamma|d_{u_\gamma}, D_{u_\gamma}), \end{aligned} \tag{19}$$

where

$$D_{u_\gamma} = v_\gamma \sum_{i=1}^{N} \gamma_i + \beta_\gamma, \quad d_{u_\gamma} = N v_\gamma + \alpha_\gamma,$$

and

$$\begin{aligned} &p(v_\gamma|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N, u_\gamma) \\ &\propto \prod_{i=1}^{N} \frac{(v_\gamma/u_\gamma)^{v_\gamma} \gamma_i^{v_\gamma-1}}{\Gamma(v_\gamma)} \exp\left(\frac{v_\gamma \gamma_i}{-u_\gamma}\right) \exp(-\xi_\gamma v_\gamma) \\ &= \frac{(v_\gamma/u_\gamma)^{N v_\gamma}}{\Gamma(v_\gamma)^N} \exp\left\{-\left[\sum_{i=1}^{N}\left(\frac{\gamma_i}{u_\gamma} - \log\gamma_i\right) + \xi_\gamma\right] v_\gamma\right\}. \end{aligned} \tag{20}$$

The condition posterior distribution of v_γ is also an unknown distribution, and takes the same form as the condition posterior distribution of v_η. Similarly, the same Metropolis step used for sampling v_η in Section 3.2.6 can be used to obtain a sample of v_γ.

In summary, (10)-(20) constitute the Metropolis-within-Gibbs sampler algorithm shown in Algorithm 1, used to draw posterior samples for each parameter in $\boldsymbol{\Psi}$ and each variable in $\{\mathbf{x}_i, i = 1, 2, \ldots, N\}$. After the constructed MCMC chain stabilizes, the simulated random samples can be used to estimate the posterior expectation $\hat{\boldsymbol{\Psi}} = \{\hat{\mu}, \hat{\kappa}, \hat{v}_\eta, \hat{u}_\eta, \hat{v}_\gamma, \hat{u}_\gamma\}$, $\{\hat{\mathbf{x}}_i, i = 1, 2, \ldots, N\}$, and their confidence intervals.

Algorithm 1 Metropolis-Within-Gibbs Sampler for Conditional Posterior of $\boldsymbol{\Psi}$

Set initial value $k = 0$, $\{\boldsymbol{\theta}_1^{(0)}, \boldsymbol{\theta}_2^{(0)}, \ldots, \boldsymbol{\theta}_N^{(0)}, \boldsymbol{\Psi}^{(0)}\}$..
for $k = 1, 2, \ldots$ **do**
  for $i = 1, \ldots, N$ **do**
    for $j = 1, \ldots, (n_i - 1)$ **do**
      Draw $x_{i,j}^{(k)} \sim p(\cdot|\mathbf{x}_{i,(j)}^{(k-1)}, \mathbf{y}_i, \boldsymbol{\theta}_i^{(k-1)})$ using (10)
    end for
    Draw $x_{i,n_i}^{(k)} \sim p(\cdot|\mathbf{x}_{i,(n_i)}^{(k-1)}, \mathbf{y}_i, \boldsymbol{\theta}_i^{(k-1)})$ using (11)
    Draw $\lambda_i^{(k)} \sim p(\cdot|\mathbf{x}_i^{(k)}, \mathbf{y}_i, \eta_i^{(k-1)}, \gamma_i^{(k-1)}, \boldsymbol{\Psi}^{(k-1)})$ using (12)
    Draw $\eta_i^{(k)} \sim p(\cdot|\mathbf{x}_i^{(k)}, \mathbf{y}_i, \lambda_i^{(k)}, \gamma_i^{(k-1)}, \boldsymbol{\Psi}^{(k-1)})$ using (13)
    Draw $\gamma_i^{(k)} \sim p(\cdot|\mathbf{x}_i^{(k)}, \mathbf{y}_i, \lambda_i^{(k)}, \eta_i^{(k)}, \boldsymbol{\Psi}^{(k-1)})$ using (14)
  end for
  Draw $\mu^{(k)} \sim p(\cdot|\boldsymbol{\theta}_1^{(k)}, \boldsymbol{\theta}_2^{(k)}, \ldots, \boldsymbol{\theta}_N^{(k)}, \kappa^{(k-1)})$ using (15)
  Draw $\kappa^{(k)} \sim p(\cdot|\boldsymbol{\theta}_1^{(k)}, \boldsymbol{\theta}_2^{(k)}, \ldots, \boldsymbol{\theta}_N^{(k)}, \mu^{(k)})$ using (16)
  Draw $v_\eta^{(k)} \sim p(\cdot|\boldsymbol{\theta}_1^{(k)}, \boldsymbol{\theta}_2^{(k)}, \ldots, \boldsymbol{\theta}_N^{(k)}, u_\eta^{(k-1)})$ using (18)
  Draw $u_\eta^{(k)} \sim p(\cdot|\boldsymbol{\theta}_1^{(k)}, \boldsymbol{\theta}_2^{(k)}, \ldots, \boldsymbol{\theta}_N^{(k)}, v_\eta^{(k)})$ using (17)
  Draw $v_\gamma^{(k)} \sim p(\cdot|\boldsymbol{\theta}_1^{(k)}, \boldsymbol{\theta}_2^{(k)}, \ldots, \boldsymbol{\theta}_N^{(k)}, u_\gamma^{(k-1)})$ using (20)
  Draw $u_\gamma^{(k)} \sim p(\cdot|\boldsymbol{\theta}_1^{(k)}, \boldsymbol{\theta}_2^{(k)}, \ldots, \boldsymbol{\theta}_N^{(k)}, v_\gamma^{(k)})$ using (19)
end for
Estimate the posterior expectations $\hat{\boldsymbol{\Psi}} = \{\hat{\mu}, \hat{\kappa}, \hat{v}_\eta, \hat{u}_\eta, \hat{v}_\gamma, \hat{u}_\gamma\}$.

IV. FAILURE TIME DISTRIBUTION

Similar to previous studies on degradation modeling [8], [10], [14], [17], we assume that a failure occurs when the degradation process $\{X(t), t \geq 0\}$ first hits a preset failure threshold level. Correspondingly, the concept of the first hitting time (FHT) is used to define the failure time. Without loss of generality, we assume that the degradation process tends to increase with time, and the failure threshold level is w. Based on the FHT concept, the failure time (lifetime) T of a system can be defined as

$$T = \inf\{t{:}X(t) \geq w|X(0) < w\}.$$

For the Wiener process degradation model defined in (1), the failure time T, given individual parameter $\{\lambda, \eta\}$ and initial state $X(0)$, follows an inverse Gaussian distribution [6]. Moreover, its PDF $f_T(t|\lambda, \eta)$ and cumulative distribution function (CDF) $F_T(t|\lambda, \eta)$ are as follows, respectively [15]

$$f_T(t|\lambda, \eta) = \frac{\tilde{w}\eta^{1/2}}{\sqrt{2\pi t^3}} \exp\left(-\frac{(\tilde{w} - \lambda t)^2 \eta}{2t}\right),$$

$$F_T(t|\lambda, \eta) = \Phi\left(\frac{(\lambda t - \tilde{w})}{\sqrt{\eta^{-1} t}}\right) + \exp\left(\frac{2\lambda\tilde{w}}{\eta^{-1}}\right) \Phi\left(-\frac{\tilde{w} + \lambda t}{\sqrt{\eta^{-1} t}}\right).$$

where $\tilde{w} = w - X(0)$.

Recall in Section 2 that λ and η are random parameters that follow the distributions defined in (6) for models M0 and M1. Using the law of total probability, the PDF and CDF of the failure time of models M0 and M1 can be respectively computed as follows [12]

$$\begin{aligned} &f_{T|\mathrm{M0,M1}}(t|\mu, \kappa, v_\eta, u_\eta) \\ &= \int\int f_T(t|\lambda, \eta) p(\lambda, \eta|\mu, \kappa, v_\eta, u_\eta) d\lambda d\eta \\ &= \frac{\Gamma(v_\eta + \frac{1}{2}) \tilde{w} u_\eta^{1/2}}{\sqrt{2\pi t^3 v_\eta(\kappa^{-1} t + 1)}\Gamma(v_\eta)} \left(1 + \frac{u_\eta(\tilde{w} - \mu t)^2}{2v_\eta(\kappa^{-1} t^2 + t)}\right)^{-\frac{v_\eta+1}{2}}, \end{aligned} \tag{21}$$

and

$$F_{T|\mathrm{M0,M1}}(t|\mu,\kappa,v_\eta,u_\eta) = \int_{-\infty}^{t} f_T(\tau|\mu,\kappa,v_\eta,u_\eta)d\tau = 1 - T_{2v_\eta}\left(-\frac{(\mu t-\tilde{w})\sqrt{u_\eta}}{\sqrt{\kappa^{-1}t^2+t}}\right), \tag{22}$$

where $T_v(\cdot)$ is the CDF of the standard t distribution with v degrees of freedom.

For model M2, only λ is the random parameter. Thus, the PDF and CDF of the failure time T of model M2 are respectively [10]

$$f_{T|\mathrm{M2}}(t|\,\mu,\kappa,\eta) = \frac{\tilde{w}}{\sqrt{2\pi t^3(\kappa^{-1}t+\eta^{-1})}}\exp\left(\frac{-(\tilde{w}-\mu t)^2}{2t(\kappa^{-1}t+\eta^{-1})}\right), \tag{23}$$

and

$$F_{T|\mathrm{M2}}(t|\,\mu,\kappa,\eta) = \Phi\left(\frac{(\mu t-\tilde{w})}{\sqrt{\kappa^{-1}t^2+\eta^{-1}t}}\right) + \exp\left(2(\mu+\kappa^{-1}\tilde{w}\eta)\tilde{w}\eta\right)\Phi\left(-\frac{\mu t+(2\kappa^{-1}\eta t+1)\tilde{w}}{\sqrt{\kappa^{-1}t^2+\eta^{-1}t}}\right), \tag{24}$$

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

It should be noted that, although models M0 and M1 have the same form of the failure time distribution, the consideration of the heteroscedastic measurement errors will affect the estimation of the model parameter $\mathbf{\Psi}$, which in turn leads to differences in the failure time prediction. This effect and differences will be illustrated in subsequent experimental studies.

V. EXPERIMENTAL STUDIES

In this section, a numerical simulation and case study are provided to verify the performance of the presented method. First, we evaluate the usefulness of Algorithm 1 with a set of simulated data. Subsequently, laser degradation data [21] are used as a practical case to compare the model fitting and prediction performance of the proposed model M0 with the reference models M1 and M2.

A. SIMULATION

In this subsection, a set of degradation data is obtained by specifying the model parameter $\mathbf{\Psi}$ and using the Monte Carlo simulation method. Since the true values of the model parameter $\mathbf{\Psi}$ and degradation states $\{\mathbf{x}_i\}$ of the degradation data are known, the purpose of this numerical example is twofold. The first is to verify the usefulness of Algorithm 1 by comparing the estimated parameters with the true values. The second is to evaluate the estimation accuracy of the three models for the degradation states.

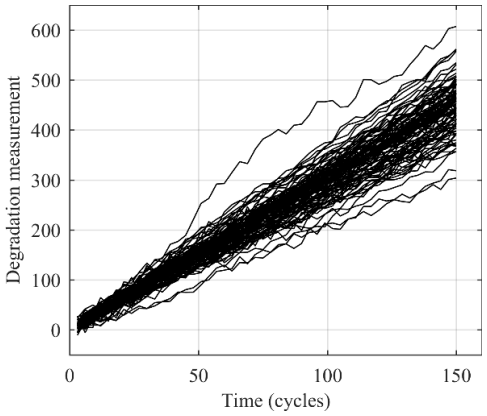

FIGURE 2. Simulated degradation paths.

In the experiment, we specify the parameter $\mathbf{\Psi}$ as

$$\mu = 3, \quad \kappa = 20, \ v_\eta = 2, \ u_\eta = 1, \ v_\gamma = 2, \ u_\gamma = 0.05,$$

and randomly generate $N = 20, 50, 100$ groups of degradation data with the same number of measurements and the same measurement time interval. Specifically, we set $n_i = 50$ and $\Delta t_{i,j} = 3$ for all i, j. Fig. 2 shows the observed degradation paths of $N = 100$ simulated data.

To verify the robustness of Algorithm 1, a weak information conjugate prior form is used to set the hyper-parameters of the model parameter $\mathbf{\Psi}$ in (8); specifically, we have

$$\omega = 0, \quad q = 0.001, \ \xi_\eta = \xi_\gamma = 0.01,$$
$$a = \alpha_\eta = \alpha_\gamma = 2, \quad b = \beta_\eta = \beta_\gamma = 0.1.$$

In the experiment, Algorithm 1 runs 30000 iterations in total. Considering the stabilization process of the Markov chains, a posterior inference is applied to $\mathbf{\Psi}$ using the remaining 10000 samples after discarding the previous 20000 transitional samples. Fig. 3 shows the diagnostic plots for the posterior samples of the parameters $\{\mu, \kappa, v_\eta, u_\eta, v_\gamma, u_\gamma\}$ in $\mathbf{\Psi}$, from which we can confirm the convergence of the Markov chains. Table 1 summarizes the results of the posterior means, standard deviations, and 95% confidence intervals of $\mathbf{\Psi}$ for $N = 20, 50, 100$ degradation data.

From Table 1, we find that under the prior setting of weak information, Algorithm 1 gives a satisfactory estimation precision, and the true value of each parameter is within the 95% confidence interval of the posterior estimation. Due to the computing-intensiveness of the MCMC method, for 100 simulated data, Algorithm 1 takes about 2045 s using MATLAB on an Intel Pentium-G4400 3.30GHz computer. For degradation modeling offline, this computing time is acceptable.

Since the actual degradation states are known in this simulation experiment, the estimation accuracy of the proposed model M0 for the degradation states can be examined. Fig. 4 shows an illustration of the estimated degradation states under models M0, M1, and M2.

As shown in Fig. 4, in the three models, the estimated degradation states from model M0 are closer to the true

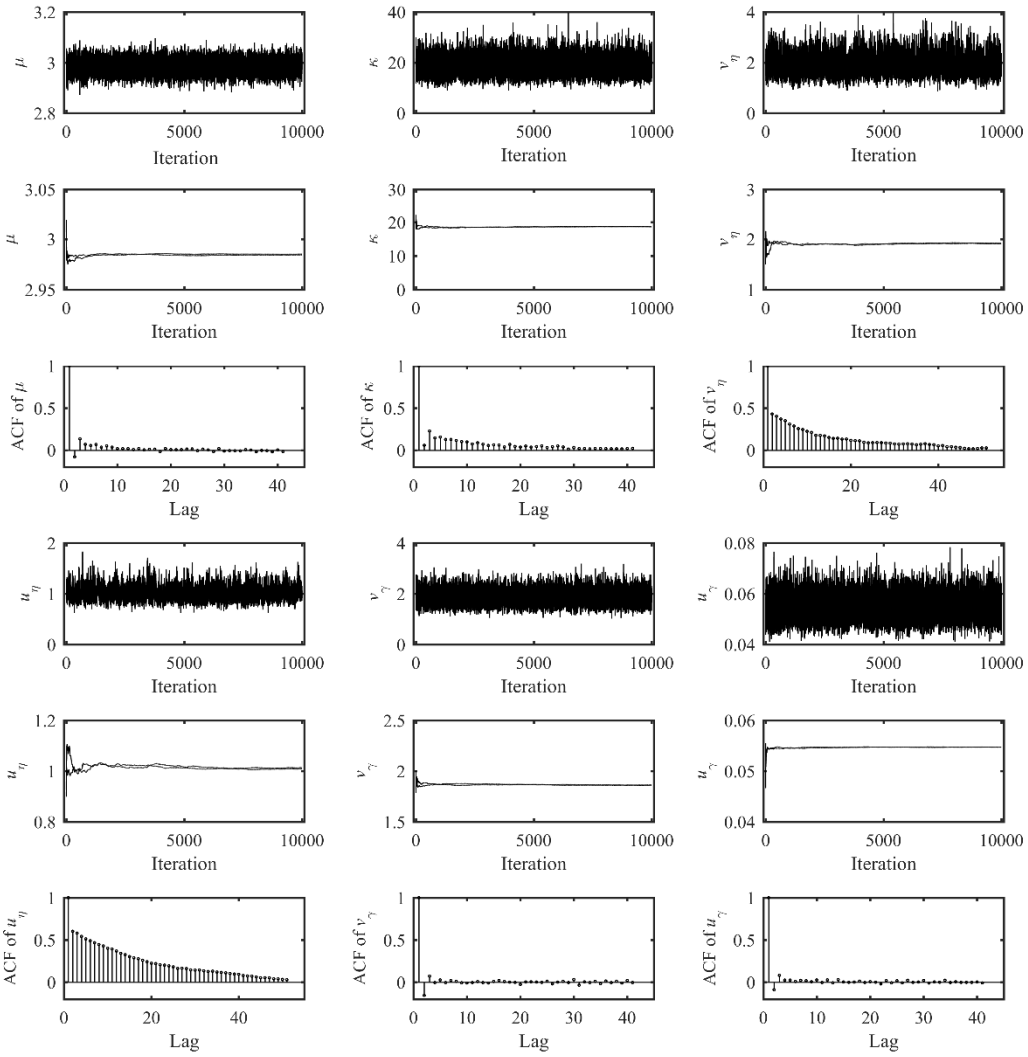

FIGURE 3. Trace, ergodic mean, and autocorrelation plots of Ψ.

degradation states for this sample path, particularly showing better robustness in cases where the measurements exhibit large deviation.

Fig. 5 shows the estimated residuals of the degradation states of all the simulation data from the three models.

It can be seen from Fig. 5 that the estimated residuals from model M0 are optimal, followed by model M1 and model M2.

Figs. 4 and 5 show that, because the proposed model M0 considers the variability in the degradation rate, degradation volatility, and error variance simultaneously, it can eliminate the influence of performance variation in degrading systems and observation instruments, and improve the estimation accuracy for the degradation states.

The mean absolute error (MAE), Akaike information criterion (AIC) [22], Bayesian information criterion (BIC) [23], and minimum description length (MDL) [24] are used as performance measures to quantify the fitting of the three models. The MAE can be calculated as

$$\mathrm{MAE} = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\left|x_{i,j} - \hat{x}_{i,j}\right| \Bigg/ \sum_{i=1}^{N} n_i,$$

the AIC can be calculated as

$$\mathrm{AIC} = -2\text{log-LF} + 2l,$$

the BIC is given by

$$\mathrm{BIC} = -2\text{log-LF} + l\ln n,$$

and MDL can be calculated as [24]

$$\mathrm{MDL} = -\text{log-LF} + (l+1)\log_2 l + \frac{l}{2}\log_2 n,$$

where $x_{i,j}$ is the actual degenerate state, $\hat{x}_{i,j}$ is the estimated posterior mean, log-LF is the log-likelihood, l is the number

TABLE 1. Estimation result of model parameter Ψ.

Parameters		μ	κ	v_η	u_η	v_γ	u_γ
True value		3	20	2	1	2	0.05
N=20	Mean	3.025	21.863	1.799	0.849	2.018	0.056
	Std.	0.068	7.461	0.821	0.247	0.676	0.014
	95% CI	[2.890, 3.155]	[10.467, 39.434]	[0.651, 3.815]	[0.508, 1.450]	[0.986, 3.608]	[0.044, 0.067]
N=50	Mean	2.968	18.324	1.775	0.929	1.845	0.057
	Std.	0.047	3.694	0.503	0.188	0.375	0.007
	95% CI	[2.872, 3.061]	[12.305, 26.739]	[1.017, 2.961]	[0.632, 1.366]	[1.203, 2.656]	[0.045, 0.071]
N=100	Mean	2.984	18.640	1.906	1.014	1.860	0.055
	Std.	0.028	3.498	0.388	0.137	0.270	0.004
	95% CI	[2.930, 3.040]	[12.636, 26.189]	[1.263, 2.768]	[0.783, 1.315]	[1.374, 2.435]	[0.046, 0.065]

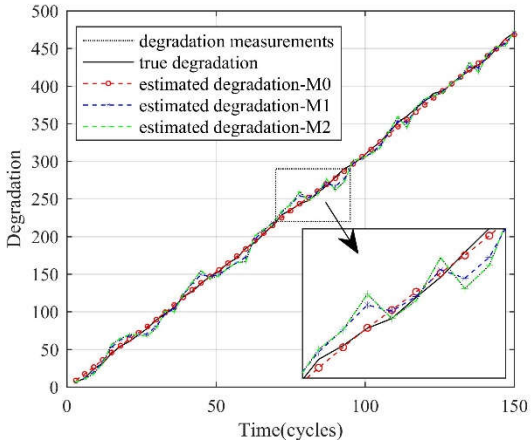

FIGURE 4. Comparison of the estimated degradation states under models M0, M1 and M2.

of model parameters, and n is the number of measurements. The three criteria AIC, BIC, and MDL prevent the over-parametric issue in modeling through a penalty term of the parameter number that helps balance the model complexity and fitting performance [23], [24]. The lower the value of the criterion, the better the model performance. Table 2 gives a comparison of the fitting of models M0, M1, and M2 for the simulation data.

Table 2 shows that model M0 has the best fitting for the simulated data among the three models. The comparison results are consistent with the simulation settings. The comparison also shows that when the observed degradation data contain heterogeneous degradation volatility and heteroscedastic measurement errors, using models M1 or M2 will induce modeling bias and reduce the fitting for the actual degradation processes.

B. CASE STUDY

Here, the degradation data of lasers [21] are used to verify the fitting and prediction accuracy of the proposed method.

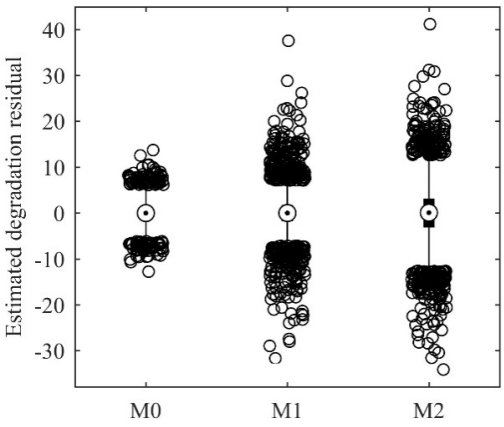

FIGURE 5. Box plots of estimated degradation residuals for all simulated data.

TABLE 2. Comparison of the fitting of three models.

Model	MAE	log-LF	AIC	BIC	MDL
M0	1.929	-9.617e3	1.925e4	1.926e4	9.652e3
M1	2.462	-1.018e4	2.037e4	2.038e4	1.021e4
M2	4.355	-2.214e4	4.429e4	4.250e4	2.126e4

The operating current increases gradually as the laser degenerates, and the laser will fail when the operating current is too high. The operating current reflects the degradation degree of the laser performance [21]. Fifteen lasers were tested at 80 °C. The percent increase in the operating current of each laser was recorded every 250 h until 4000 h. Fig. 6 shows the observed degradation paths of all the lasers. In [21], it was assumed that the failure threshold is 10 (i.e., the laser fails when the operating current increases by 10%), and the pseudo failure times of 15 lasers were obtained using a linear degradation path model. In this experiment, we also use the same failure threshold, i.e., $w = 10$. Moreover, the empirical failure distribution of the pseudo failure times and the pointwise 95% confidence intervals obtained through the Kaplan–Meier method [25] are used as the experimental reference.

With the degradation data, models M0, M1, and M2 are used to fit the degradation processes of the lasers. Table 3 lists the posterior means and standard deviations of the model parameters, as well as the log-likelihood, AIC, BIC, and MDL values.

Comparing log-LF and these performance measures, we find that model M0 outperforms models M1 and M2 for the laser degradation data. Moreover, the estimations of the parameters $\{\mu, \kappa, v_\eta, u_\eta\}$ of models M0 and M1 are significantly different. This indicates that the consideration of the heteroscedastic measurement errors will have a significant impact on degradation modeling. This impact will in turn lead to performance differences in the case of failure prediction using (21) and (22).

Fig. 7 shows the posterior means of the individual parameters $\{\lambda_i, \eta_i, \gamma_i\}$, $i = 1, 2, \ldots, 15$, of each laser. As shown,

TABLE 3. Comparison of three degradation models.

Model	Estimation	μ	κ	η	v_η	u_η	γ	v_γ	u_γ	log-LF	AIC	BIC	MDL
M0	Mean	1.891	9.830	–	1.594	12.625	–	1.617	166.948	85.305	-158.611	-147.138	-50.279
	Std.	0.055	5.598	–	0.608	3.657	–	0.780	66.407				
M1	Mean	2.015	0.279	–	29.392	23.589	38.879	–	–	70.425	-130.851	-121.290	-42.384
	Std.	0.124	0.134	–	179.904	6.499	4.715	–	–				
M2	Mean	2.038	3.318	0.755	–	–	21.717	–	–	-124.186	256.372	256.372	145.474
	Std.	0.221	1.518	0.285	–	–	4.200	–	–				

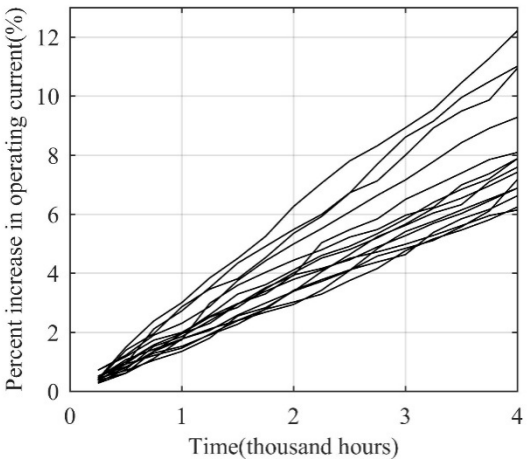

FIGURE 6. Observed degradation paths of lasers.

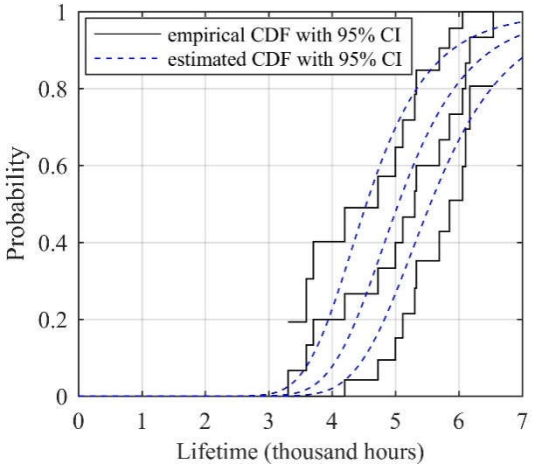

FIGURE 8. Comparison between the estimated CDF using model M0 and the empirical CDF.

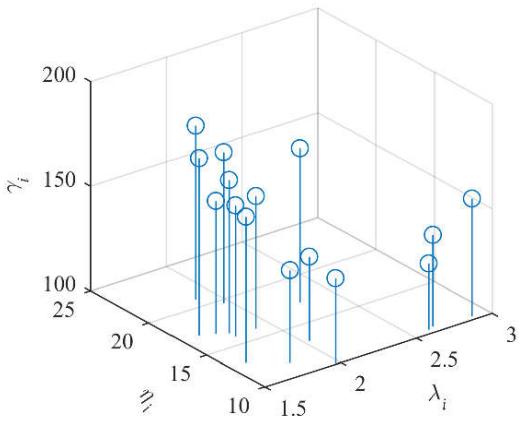

FIGURE 7. Scatter plot of posterior mean for all individual parameters.

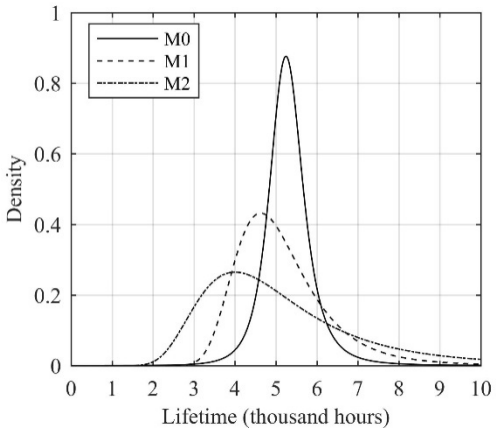

FIGURE 9. Comparison of the estimated PDFs under models M0, M1, and M2.

the degradation process of each laser is evidently different in terms of the degradation rate, degradation volatility, and error variance. These differences indicate that the influences of the multi-source heterogeneity should not be ignored in the degradation modeling of the laser data.

The empirical CDF of the failure time is compared with the estimated CDF obtained using the proposed model M0, as shown in Fig. 8, including the 95% confidence intervals. This figure shows that the estimated failure time distribution agrees well with the empirical distribution, indicating goodness-of-fit. The PDF of the failure time obtained by models M0, M1, and M2 are compared as shown in Fig. 9.

As shown in Fig. 9, the PDF of the failure time estimated by model M0 is narrower than the ones estimated by models

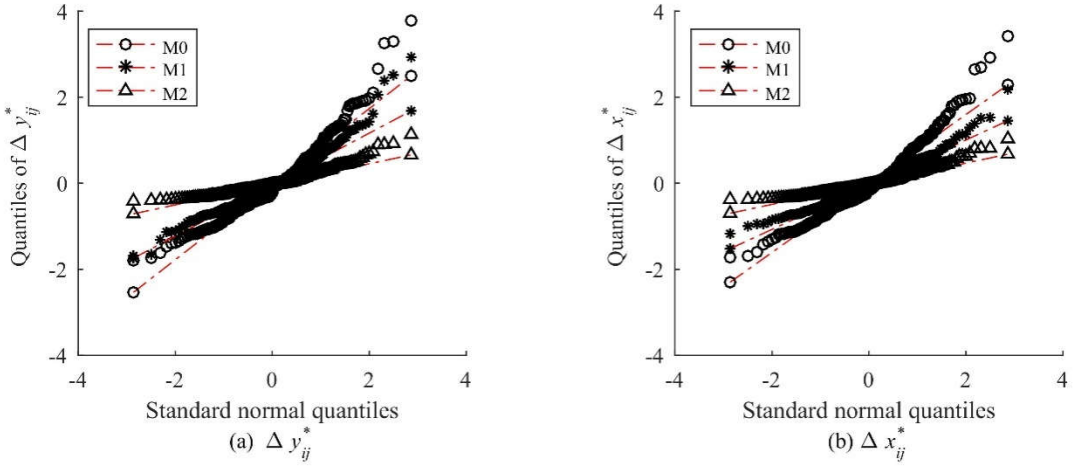

FIGURE 10. Q-Q plots of $\Delta x_{i,j}^*$, $\Delta y_{i,j}^*$ for M0, M1, and M2.

M1 and M2, and its peak value is closer to the average failure time of 5109.7 h given in [21]. This indicates that the prediction uncertainty of model M0 is minimal and that the prediction accuracy of M0 is optimal.

While the actual degradation states of the lasers are unknown, the degradation state estimation performance of models M0, M1, and M2 can be tested by analyzing the consistency between the estimated states and the modeling assumptions. From (3), we see that the normalized increment $\Delta x_{i,j}^*$ of the degradation states and the normalized residual $\Delta y_{i,j}^*$ of the degradation measurements should follow a standard normal distribution. $\Delta x_{i,j}^*$ and $\Delta y_{i,j}^*$ can be calculated as follows:

$$\Delta x_{i,j}^* = \frac{\Delta \hat{x}_{i,j} - \hat{\lambda}_i \Delta t_{i,j}}{(\hat{\eta}_i^{-1} \Delta t_{i,j})^{1/2}}, \quad \Delta y_{i,j}^* = \frac{y_{i,j} - \hat{x}_{i,j}}{\hat{\gamma}_i^{-1/2}}, \tag{25}$$

where $\hat{x}_{i,j}$ and $\{\hat{\lambda}_i, \hat{\eta}_i, \hat{\gamma}_i\}$ are the estimated degradation state and the individual parameters, respectively; $\Delta \hat{x}_{i,j} = \hat{x}_{i,j} - \hat{x}_{i,j-1}$, $i = 1, 2, \ldots, 15, j = 1, 2, \ldots, n_i$.

Based on the estimated states of models M0, M1, M2, and (25), the Q-Q plots of $\Delta x_{i,j}^*$ and $\Delta y_{i,j}^*$ of the three models are shown in Fig. 10.

Fig. 10 shows that the Q-Q plots of models M0, M1, and M2 are close to different straight lines. Thus, the assumptions that $\Delta x_{i,j}^*$ and $\Delta y_{i,j}^*$ of all three models are normally distributed cannot be rejected; nevertheless, there are differences in the extent to which $\Delta x_{i,j}^*$ and $\Delta y_{i,j}^*$ of the three models conform to the standard normal distribution. The statistical characteristics of $\Delta x_{i,j}^*$ and $\Delta y_{i,j}^*$ are further compared in Table 4.

As listed in Table 4, when the assumptions of the normal distribution are not rejected, $\Delta x_{i,j}^*$ and $\Delta y_{i,j}^*$ from model M0 are closer to the standard normal distribution, i.e., M0 has the best consistency with the modeling assumptions, followed by M1 and M2. Table 4 also shows that model M0 can eliminate the inter-individual variation and improve the estimation accuracy for the actual degradation processes.

TABLE 4. Comparison of the statistical characteristics of $\Delta x_{i,j}^*$, $\Delta y_{i,j}^*$.

Model	$\Delta y_{i,j}^*$		$\Delta x_{i,j}^*$	
	Mean	Std.	Mean	Std.
M0	0.001	0.886	0.002	0.866
M1	-0.002	0.679	0.002	0.528
M2	-0.001	0.271	0.001	0.253

The comparison results of the consistency of the modeling assumptions in Table 4 are in line with those of the AIC measures in Table 3 and the failure probability density in Fig. 9. This further indicates that model M0 has better fitting performance and failure prediction accuracy than models M1 and M2 in the laser degradation modeling.

VI. CONCLUSION

Because of the complexity and randomness of actual degradation failure processes, the degradation data from different individuals in a system population exhibit significant heterogeneity. To improve the fitting ability of degradation models, the heterogeneity due to the inherent differences in the degrading systems and measuring instruments, as well as due to the extrinsic environmental differences, should not be neglected. This paper proposes a random effects Wiener process model with heteroscedastic measurement errors to capture the multi-source heterogeneity. The results of a numerical example and practical case study verified the feasibility and effectiveness of the proposed method. The main conclusions are as follows:

1) The proposed model, with the drift, diffusion, and error variance taken as unit-specific random effects, can further interpret and quantify the influence of multi-source heterogeneity. It is more reasonable and generalizable than models that consider only partial degradation heterogeneity factors.

2) Based on a hierarchical prior structure, a Metropolis-within-Gibbs algorithm was adopted to estimate the posterior distributions of the actual degradation states and the model parameters. The proposed algorithm can efficiently deal with the statistical inference of complex models with multiple random parameters. The simulation results helped verify the validity of the algorithm.

3) The experimental results of laser degradation data show that the interpretation and quantification of multiple heterogeneity factors and their combined effects in degradation modeling can eliminate the inter-individual variation, and further improve the model fitting and prediction accuracy.

4) In the future, considering the more general time-varying degradation rate in practice, it is necessary to extend the work conducted in this study to a nonlinear degradation model considering multi-source heterogeneity.

REFERENCES

[1] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, Oct. 2006.
[2] 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, Aug. 2011.
[3] 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, Jan. 2015.
[4] Z. Zhang, X. Si, C. Hu, and X. Kong, "Degradation modeling–based remaining useful life estimation: A review on approaches for systems with heterogeneity," *Proc. Inst. Mech. Eng. O, J. Risk Rel.*, vol. 229, no. 4, pp. 343–355, Aug. 2015.
[5] 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.
[6] 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.
[7] 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, no. 6, pp. 543–557, Jun. 2005.
[8] X.-S. Si, "An adaptive prognostic approach via nonlinear degradation modeling: Application to battery data," *IEEE Trans. Ind. Electron.*, vol. 62, no. 8, pp. 5082–5096, Aug. 2015.
[9] G. A. Whitmore and F. Schenkelberg, "Modelling accelerated degradation data using Wiener diffusion with a time scale transformation," *Lifetime Data Anal.*, vol. 3, no. 1, pp. 27–45, 1997.
[10] G. Jin, D. E. Matthews, and Z. Zhou, "A Bayesian framework for on-line degradation assessment and residual life prediction of secondary batteries inspacecraft," *Rel. Eng. Syst. Saf.*, vol. 113, pp. 7–20, May 2013.
[11] Z.-S. Ye, N. Chen, and Y. Shen, "A new class of Wiener process models for degradation analysis," *Rel. Eng. Syst. Saf.*, vol. 139, pp. 58–67, Jul. 2015.
[12] X. Wang, "Wiener processes with random effects for degradation data," *J. Multivariate Anal.*, vol. 101, no. 2, pp. 340–351, Feb. 2010.
[13] X. Wang, P. Jiang, B. Guo, and Z. Cheng, "Real-time reliability evaluation with a general Wiener process-based degradation model," *Qual. Rel. Eng. Int.*, vol. 30, no. 2, pp. 205–220, Mar. 2014.
[14] H. Wang, X. Ma, and Y. Zhao, "An improved Wiener process model with adaptive drift and diffusion for online remaining useful life prediction," *Mech. Syst. Signal Process.*, vol. 127, pp. 370–387, Jul. 2019.
[15] 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. 2014.
[16] 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.
[17] 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.
[18] K. P. Murphy, "Conjugate Bayesian analysis of the Gaussian distribution," Univ. Brit. Columbia, Vancouver, BC, Canada, Tech. Rep., 2007, p. 6. [Online]. Available: https://www.cs.ubc.ca/~murphyk/Papers/bayesGauss.pdf
[19] M. Wiper, D. R. Insua, and F. Ruggeri, "Mixtures of gamma distributions with applications," *J. Comput. Graph. Statist.*, vol. 10, no. 3, pp. 440–454, Sep. 2001.
[20] C. Andrieu, N. De Freitas, A. Doucet, and M. I. Jordan, "An introduction to MCMC for machine learning," *Mach. Learn.*, vol. 50, no. 1, pp. 5–43, Jan. 2003.
[21] W. Q. Meeker and L. A. Escobar, *Statistical Methods for Reliability Data*. New York, NY, USA: Wiley, 1998, p. 642.
[22] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.
[23] L. Wasserman, "Bayesian model selection and model averaging," *J. Math. Psychol.*, vol. 44, no. 1, pp. 92–107, Mar. 2000.
[24] A. Asensio Ramos, "The minimum description length principle and model selection in spectropolarimetry," *Astrophysical J.*, vol. 646, no. 2, pp. 1445–1451, Aug. 2006.
[25] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *J. Amer. Stat. Assoc.*, vol. 53, no. 282, pp. 457–481, Jun. 1958.

MENG XIAO received the M.Eng. degree in traffic information engineering and control from Lanzhou Jiaotong University, Lanzhou, China, in 1999. He is currently an Associate Professor with the School of Railway Tracks and Transportation, Wuyi University. His research interests include reliability engineering, degradation analysis, predictive maintenance, and application of statistics techniques.

YOUPENG ZHANG received the M.Eng. degree in traffic information engineering and control from Lanzhou Jiaotong University, Lanzhou, China, in 1991. He is currently a Professor with the School of Automation and Electrical Engineering, Lanzhou Jiaotong University. His research interests include reliability engineering, predictive maintenance, and track circuit fault detection and diagnosis.

YONGJIAN LI received the M.Eng. and Ph.D. degrees in mechanical engineering from Southwest Jiaotong University, Chengdu, China, in 2013 and 2017, respectively. He is currently a Lecturer with the School of Railway Tracks and Transportation, Wuyi University, China. His research interests include signal processing and data mining for machine health monitoring and fault diagnosis.

WENXIAN WANG received the M.Eng. and Ph.D. degrees in traffic and transportation planning and management from Southwest Jiaotong University, Chengdu, China, in 2012 and 2017, respectively. He is currently a Lecturer with the School of Railway Tracks and Transportation, Wuyi University, China. His research interests include optimization of passenger transport organization and intelligent optimization algorithm.