A Data-Driven Modeling Method for Stochastic Nonlinear Degradation Process With Application to RUL Estimation

Yuhan Zhang, Ying Yang, *Senior Member, IEEE*, He Li, Xianchao Xiu, *Member, IEEE*, and Wanquan Liu, *Senior Member, IEEE*

***Abstract*—This article proposes a novel modeling method for the stochastic nonlinear degradation process by using the relevance vector machine (RVM), which can describe the nonlinearity of degradation process more flexibly and accurately. Compared with the existing methods, where degradation processes are modeled as the Wiener process with a nonlinear drift function formulized as the power law or exponential law, this kind of modeling method can characterize degradation processes with more nonlinear behavior. Instead of modeling the drift coefficient of the Wiener process directly, the weighted combination of basis functions is utilized to express the increment of the Wiener process and the parameters are calculated by a sparse Bayesian learning algorithm. Based on the proposed model, a numerical approximation formula for the probability density function (PDF) of the remaining useful life (RUL) is derived. Finally, comparison studies, including a numerical simulation and a practical case, are provided to demonstrate the effectiveness and the accuracy of the proposed methods for RUL estimation.**

***Index Terms*—Data driven, degradation process, remaining useful life (RUL).**

Acronyms

AIC	Akaike information criterion.
BM	Brownian motion.
CM	Condition monitoring.
FHT	First hitting time.
MLE	Maximum likelihood estimation.
MSE	Mean squared error.
PDF	Probability density function.
PHM	Prognostics and health management.
RMS	Root mean square.
RUL	Remaining useful life.
RVM	Relevance vector machine.
SVM	Support vector machine.
WP	Wiener process.

Manuscript received November 1, 2019; revised March 25, 2020, August 5, 2020, and January 10, 2021; accepted April 3, 2021. Date of publication April 29, 2021; date of current version May 18, 2022. This work was supported in part by the National Natural Science Foundation (NNSF) of China under Grant 61633001, Grant 12001019, and Grant U1713223; and in part by the China's Ministry of Industry and Information Technology under the Research and Development Project of Intelligent Ship under Grant 1.0 ([2016]544). This article was recommended by Associate Editor Z. Wang.
(Corresponding author: Ying Yang.)

Yuhan Zhang, Ying Yang, He Li, and Xianchao Xiu are with the State Key Laboratory for Turbulence and Complex Systems, Department of Mechanics and Engineering Science, College of Engineering, Peking University, Beijing 100871, China (e-mail: yuhanzhang@pku.edu.cn; yy@pku.edu.cn).

Wanquan Liu is with the School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou 510275, China (e-mail: wanquanliu@qq.com).
Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSMC.2021.3073052.

Digital Object Identifier 10.1109/TSMC.2021.3073052

Notation

$B(\cdot)$	Standard BM.
X(t)	Degradation state at time t.
$\mu(t;\boldsymbol{\theta})$	Drift coefficient of $X(t)$.
σ	Diffusion coefficient of $X(t)$.
$K(\cdot)$	Basis function of RVM model.
τ	Step length of degradation increment.
$\int_t^{t+\tau}\mu(t;\boldsymbol{\theta})dt$	Drift increment of τ step length.
ξ	Threshold of degradation process.
T	Lifetime.
L_k	Random variable of the RUL at time t_k.
$\boldsymbol{\theta}$	Parameter vector in $\mu(t,\boldsymbol{\theta})$.
$f_T(k_T\tau)$	PDF of the FHT at time t_k.
$f_L(k_L\tau)$	PDF of the RUL at time t_k.

I. Introduction

PROGNOSTICS and health management (PHM) based on condition monitoring (CM) has been studied intensively over the past decades, due to its contributions to working availability improvement, operating risks, and costs reduction for military and commercial systems [1]. The remaining useful life (RUL) estimation, an important part of PHM, can be scheduled to plan replacement activities and arrange a maintenance plan to ensure the deteriorating systems operate reliably, safely, and economically [2], [3]. Lithium-ion batteries [4], rotating bearings [5]–[7], a power transformer [8], circuit breakers [9], or a pressurized water reactor [10] can be found in the literature as examples. It is the uncertainties in the working environments of the system, the variability of the individual system in a population, and the random errors in measurements that make stochastic dynamics become the commonest feature in the actual degradation processes. Since stochastic models possess a considerable skill in describing stochastic dynamics during degradation, degradation analysis and RUL estimation based on stochastic models are favored by many researchers [2]. The representative stochastic models include the Wiener process [7], [11]–[13], Gamma process [14], [15] and inverse Gaussian process [16], [17]. The Wiener process is appropriate for nonmonotonous degradation processes modeling in comparison to the Gamma process and the inverse Gaussian process which apply only to modeling degradation processes with monotonous paths. For that reason, more recent attention has been paid to this model because it can depict the system's degradation signals nicely and flexibly, including laser device, gyroscope, aluminum alloy fatigue crack [13], milling machine [18], blast furnace [19], and battery [20].

In fact, stochasticity and nonlinearity are two crucial elements in the degradation processes of complicated systems, which are more likely to make the degradation process accelerate later in life. Thus, for degradation modeling, how to deal with the stochasticity and nonlinearity in the degradation process is a problem that has attracted much attention in recent years. The state transformations [21] or time-scale transformations [22] to linearize the nonlinear degradation process to the linear Wiener degradation model was presented as a possible solution. According to [13], the nonlinear diffusion process-based model $dX(t) = \mu(t; \boldsymbol{\theta})dt + \sigma(t; \boldsymbol{\theta})dB(t)$ can describe the nonlinearity and stochasticity of the degradation. Based on it, Si *et al.* [13] provided the analytical approximation to the distributions of the first hitting time (FHT) by the diffusion process crossing a threshold level and the RUL in closed forms. Subsequently, additional extensions and applications [23]–[25] of the degradation process on this basis were carried out. Among all the types of degradation models used in existing studies, the power-law model with $\mu(t; \boldsymbol{\theta}) = ab \cdot t^{b-1}$ and the exponential-law model with $\mu(t; \boldsymbol{\theta}) = ab \cdot \exp(bt)$ gain the most popularity in the scientific community when nonlinear degradation processes were modeled. However, the two models will struggle to harness the degradation cases with more complicated nonlinear behavior. Zhang *et al.* [24] models the drift coefficient with a linear combination of simple functions, which can be regarded as a representation of general nonlinearity [2]. But, this approach mostly focuses on the random effect modeling in the degradation process, and the simulation part only gives an example of the linear Wiener process. In addition, this method needs to determine the explicit forms of these simple functions as well as the number of parameters clearly at the beginning of modeling. This is extremely difficult with a lack of particular prior knowledge about the nature of degradation. A possible way to model the nonlinearity of the stochastic degradation process is the relevance vector machine (RVM) based on statistical learning. RVM proposed by Tipping [26] is a probabilistic model with the same functional form as the support vector machine (SVM), which offers a number of additional advantages, including the superiority of probability prediction, the automatic identification of parameters, and the convenience of using arbitrary basis functions.

In accordance with the promising results of an RVM, this article proposes a learning-based data-driven modeling method to represent the degradation process with more nonlinear behavior in Section II. In the proposed model, the drift increment of the Wiener-process-based degradation is modeled as a weighted sum of basis functions under a Bayesian framework, namely, the RVM model, which provides a more flexible and general description of the nonlinear degradation process. In this article, the Bayesian learning algorithm is adopted to identify the RVM prediction model of the drift increment of the Wiener-process-based degradation. Moreover, by identifying and predicting the increment with some certain nonlinear basis functions, the nonlinear representation of the degradation process can be derived equivalently. Besides, due to the excellent prediction performance of the RVM, the degradation process with more nonlinear behavior can be reasonably modeled when the appropriate basis function is selected. Then, based on the theorems in [13] and parameters estimated by the Bayesian learning algorithm, two main theorems for computing the numerical approximate probability density function (PDF) values of the FHT and RUL of degradation process are presented. Following them, an RUL estimation algorithm is developed in Section IV.

The main contribution of this article can be summarized as follows.

1) A novel and data-driven modeling and RUL estimation approach for degradation processes with more nonlinear behavior is put forward. In comparison with the existing methods for the nonlinear degradation process, the nonlinearity in the degradation process can be described more accurately and flexibly, which will be proven in Section II.
2) The degradation process is modeled by representing the degradation drift increment by the weighted sum of basis functions, namely, the RVM model. When some certain basis functions are selected, linear degradation processes and nonlinear degradation processes with power law and exponential law can be included as particular cases.
3) The RVM model is utilized to depict the stochastic nonlinear degradation process, in a way that a part of the degradation data becomes directly parameters in the model. The unused part helps to modify and improve the model.
4) An approximation formula for the PDF of the RUL is presented. Comparison studies, including a numerical simulation and a practical degraded bearing case, demonstrate that the presented approach offers better performance than the existing state-of-the-art methods [5], [13] for RUL estimation of nonlinear degradation process.

This article is structured as follows. In Section II, two basic nonlinear degradation process models are reviewed. In addition, a new nonlinear degradation process model based on RVM is provided. Section III presents the parameter calculation method under a Bayesian framework. The primary theoretical results for computing the numerical approximate PDF values of the FHT and the RUL, as well as the RUL estimation algorithm, are proposed in Section IV. In Section V, a numerical simulation and a practical case are presented to verify the proposed method. Finally, some conclusions are discussed in Section VII.

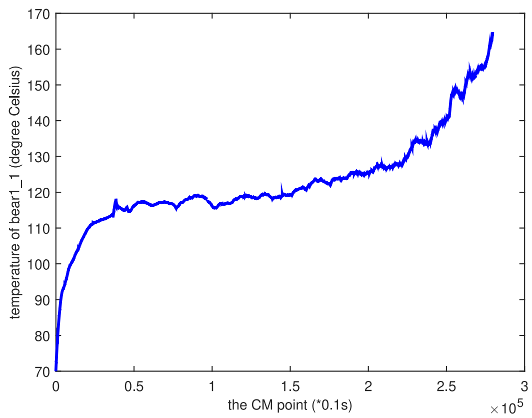


Fig. 1. Temperature measurement data of bearing1_1.

II. Nonlinear Degradation Process Model

In order to achieve nonlinear degradation modeling, a general nonlinear degradation model on the basis of the following diffusion process presented in [13] has gained many scientists' attention [19], [20], [24]. $X(t)$ denotes the degradation at time t. Then

$$dX(t) = \mu(t; \boldsymbol{\theta})dt + \sigma(t; \boldsymbol{\theta})dB(t) \tag{1}$$

where the drift coefficient function $\mu(t; \boldsymbol{\theta})$ and the diffusion coefficient function $\sigma(t; \boldsymbol{\theta})$ are both time-dependent, $\boldsymbol{\theta}$ is the parameter vector, and $B(t)$ is a standard Brownian motion (BM). Among the degradation models based on (1), rather than the diffusion coefficient function, much more attention has been paid to the drift coefficient function in the academic and engineering fields. $\sigma(t; \boldsymbol{\theta})$ is mostly fixed as a constant to partially characterize the time-varying dynamics of the degradation process, while $\mu(t; \boldsymbol{\theta})$ is formulated as a nonlinear function with respect to t to depict the nonlinearity of the stochastic degradation process. In all of these degradation models, the power-law model with

$$\mu(t; \boldsymbol{\theta}) = ab \cdot t^{b-1} \tag{2}$$

and the exponential law model with

$$\mu(t; \boldsymbol{\theta}) = ab \cdot \exp(bt) \tag{3}$$

have gained the most popularity. Relying on them, many parameter identification algorithms and RUL estimation methods have been developed extensively [19], [20], [24]. Though these models may be capable of describing a majority of common nonlinear stochastic degradation processes, there might be less suitable for dealing with the nonlinearity in some more complex cases, such as the degradation process of bearing1_1 in [27] based on the temperature measurement data in Fig. 1.

Instead of modeling the degradation process stiffly with some empirical forms, here, a method is proposed to overcome the lack of priori knowledge over the drift coefficient function form. The new method shifts the focus from the degradation to the increment of it. The basic idea is to characterize the degradation increment using an RVM model. Given a stochastic degradation process formulized as follows:

$$X(t) = X(0) + \int_0^t \mu(t; \boldsymbol{\theta})dt + \sigma B(t) \tag{4}$$

where the diffusion coefficient σ is constant, the degradation increment of τ step length is given by

$$\Delta X_\tau(t) = \int_t^{t+\tau} \mu(t; \boldsymbol{\theta})dt + \eta(t) \tag{5}$$

where the diffusion increment $\eta(t) = \sigma B(t+\tau) - \sigma B(t) \sim N(0, \sigma^2\tau)$ obeys a normal distribution and the drift increment function $\int_t^{t+\tau} \mu(t; \boldsymbol{\theta})dt$ has the form

$$\int_t^{t+\tau} \mu(t; \boldsymbol{\theta})dt = \sum_{i=1}^{m} \omega_i K(t, t_i) + \omega_0 \tag{6}$$

where ω_i is the weight of the basis function $K(t, t_i)$, ω_0 is the bias term, m is the sample size of the RVM model dataset, the step length $\tau = t_{i+1} - t_i$, and $K(t, t_i)$ can have different types of forms as stated as follows.

1) *Linear Basis Function:*

$$K(t, t_i) = t_i t + 1. \tag{7}$$

If $t_i = 0$, then $\mu(t; \boldsymbol{\theta})$ degenerates into constant μ, and (4) becomes the conventional linear drifted model, like in [28] and [29].

2) *Polynomial Basis Function:*

$$K(t, t_i) = (t - t_i)^p. \tag{8}$$

If $p = b$, then $\mu(t; \boldsymbol{\theta})$ can degenerate into (2), and (4) becomes the power-law model. Under the circumstances, the degradation process based on (3) turns into a particular case of the degradation process based on (8).

3) *Gaussian Basis Function:*

$$K(t, t_i) = \exp\left(-\frac{(t - t_i)^2}{q^2}\right) \tag{9}$$

which is a commonly used basis function in nonlinear model prediction.

4) *Exponential Basis Function:*

$$K(t, t_i) = \exp\left(\frac{(t - t_i)}{q}\right). \tag{10}$$

The degradation process based on (10) can include the degradation process in (3) as a particular case.

The data-driven modeling method refers to that the degradation increment data, which drives to select the number of parameters and the form of the drift coefficient function rather than using a single power-law or exponential law empirical model to describe the stochastic nonlinear degradation process. The parameters $\{t_i\}$ in the model are all from degradation data, while the weights $\{\omega_i\}$ and the diffusion coefficient σ are adjusted based on degradation increment data.

III. Parameter Calculation Algorithm

In this section, a sparse Bayesian parameter calculation algorithm [26] is adopted for the nonlinear degradation model. The parameter vector $\boldsymbol{\theta}$ in the model mentioned above consists of two parts, $\boldsymbol{\theta} = [\boldsymbol{\theta_0}, \boldsymbol{\theta_{RVM}}]^T$. The parameters in $\boldsymbol{\theta_{RVM}} = [\omega_0, \omega_1, \ldots, \omega_m, t_1, \ldots, t_m]^T$ and σ are calculated with the help of a sparse Bayesian learning algorithm. The basic parameters of the basis function $\boldsymbol{\theta_0} = p$ in (8), as well as $\boldsymbol{\theta_0} = q$ in (9) and (10), are the basic parameters of the basis function determined by cross validation.

In the following part, it is discussed, how to compute the parameters in $\boldsymbol{\theta_{\mathrm{RVM}}}$ and σ with the degradation increment data. It is worth noting that the sparse Bayesian learning algorithm proposed by Tipping [26] is valid for $\boldsymbol{\theta_{\mathrm{RVM}}}$ and σ calculation in the proposed degradation model because the property of the increment of the Wiener-process-based degradation model fits the basic assumption of the sparse Bayesian learning algorithm. On the one hand, the RVM model assumes that the targets are samples from the model with additive noise

$$\mathrm{tar}_n = f(t_n; \boldsymbol{\omega}) + \varepsilon_n \tag{11}$$

where $\boldsymbol{\omega} = [\omega_0, \omega_1, \ldots, \omega_n]^T$ are weights and ε_n are independent and assumed to be Gaussian with a mean zero and the variance ϵ^2. Hence, the distribution $p(\mathrm{tar}_n|\boldsymbol{t}) = N(\mathrm{tar}_n|f(t_n;\boldsymbol{\omega}), \epsilon^2)$ refers to a normal distribution over tar_n with a mean $f(t_n;\boldsymbol{\omega})$ and the variance ϵ^2 [26]. On the other hand, according to the definition of the Wiener process [26], the BM has stationary and independent increments. $\{\Delta X_\tau(t_k)\}_{k=1}^m$ are independent and the distribution of $\Delta X_\tau(t_k)$ depends on t_k only through the difference τ; which is normal with a mean $\int_t^{t+\tau} \mu(t;\boldsymbol{\theta})dt$ and the variance $\sigma^2\tau$ [30]. The independence and the gaussian distribution of the degradation increment data are consistent with the premise of RVM modeling, comparing (5) and (11). The calculated variance of errors ϵ^2 in the RVM model corresponds to the calculation of $\sigma^2\tau$ in the proposed model. For that reason, the sparse Bayesian learning algorithm is a proper tool to calculate the parameters $\boldsymbol{\theta_{\mathrm{RVM}}}$ and σ in the degradation model, which will be described in detail next.

Given a degradation increment dataset of input–target pairs $\{t_k, \Delta X_\tau(t_k)\}_{k=1}^m$, $\Delta X_\tau(t_k) \in R$, $t_k \in R$, the output of the RVM model $\int_t^{t+\tau} \mu(t;\boldsymbol{\theta})dt$ is

$$\begin{aligned}\int_t^{t+\tau} \mu(t;\boldsymbol{\theta})dt &= f(t,\omega) \\ &= \sum_{i=1}^{m} \omega_i K(t, t_i) + \omega_0.\end{aligned} \tag{12}$$

Following the assumption of independence for $\Delta X_\tau(t_k)$, the likelihood of the complete dataset can be written as:

$$p\Big(\boldsymbol{\Delta X}|\boldsymbol{\omega}, \sigma^2\tau\Big) = \Big(2\pi\sigma^2\tau\Big)^{-m/2}\exp\left(-\frac{\|\boldsymbol{\Delta X} - \boldsymbol{\Phi\omega}\|^2}{2\sigma^2\tau}\right) \tag{13}$$

where $\boldsymbol{\Delta X} = [\Delta X_\tau(t_1), \ldots, \Delta X_\tau(t_m)]^T$, $\boldsymbol{\omega} = [\omega_0, \omega_1, \ldots, \omega_m]^T$, and $\boldsymbol{\Phi}$ is an $m \times (m+1)$ design matrix with $\boldsymbol{\Phi} = [\phi(t_1), \phi(t_2), \ldots, \phi(t_m)]^T$ and $\phi(t_i) = [1, K(t_i, t_1), \ldots, K(t_i, t_m)]$.

The direct maximum likelihood estimation (MLE) of $\boldsymbol{\omega}$ and σ from (13) may result in overfitting [26]. In order to avoid this, additional measurements are required to set additional constraints on the parameters from a Bayesian point of view by explicitly defining a zero-mean Gaussian prior probability distribution over $\boldsymbol{\omega}$

$$p(\boldsymbol{\omega}|\boldsymbol{\alpha}) = \prod_{i=0}^{m} \frac{\alpha_i}{\sqrt{2\pi}}\exp\left(-\frac{\omega_i^2\alpha_i^2}{2}\right) \tag{14}$$

with $\boldsymbol{\alpha}$ as the vector of $m+1$ hyperparameters, namely, $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \ldots, \alpha_m]^T$.

Hence, the posterior distribution over ω is determined by

$$\begin{aligned}p\Big(\boldsymbol{\omega}|\boldsymbol{\Delta X}, \boldsymbol{\alpha}, \sigma^2\tau\Big) &= (2\pi)^{-\frac{m+1}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} \\ &\quad \exp\left\{-\frac{(\boldsymbol{\omega}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\omega}-\boldsymbol{\mu})}{2}\right\}\end{aligned} \tag{15}$$

where the posterior mean and covariance are, respectively

$$\boldsymbol{\mu} = \sigma^{-2}\tau^{-1}\boldsymbol{\Sigma\Phi}^T\boldsymbol{\Delta X} \tag{16}$$

$$\boldsymbol{\Sigma} = \Big(\sigma^{-2}\tau^{-1}\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \boldsymbol{A}\Big)^{-1} \tag{17}$$

where $\boldsymbol{A} = \mathrm{diag}(\alpha_0, \alpha_1, \ldots, \alpha_m)$.

By maximizing the "evidence for the hyperparameters," i.e., the marginal likelihood, the optimal values of α and σ can be derived. The marginal likelihood is obtained by

$$p\Big(\boldsymbol{\Delta X}|\boldsymbol{\alpha}, \sigma^2\tau\Big) = \int p\Big(\boldsymbol{\Delta X}|\boldsymbol{\omega}, \sigma^2\tau\Big)\cdot p(\boldsymbol{\omega}|\boldsymbol{\alpha})d\boldsymbol{\omega} \tag{18}$$

which can be calculated by

$$\begin{aligned}p\Big(\boldsymbol{\Delta X}|\boldsymbol{\alpha}, \sigma^2\tau\Big) &= (2\pi)^{-\frac{m}{2}}\cdot\Big|\sigma^2\tau\boldsymbol{I} + \boldsymbol{\Phi A}^{-1}\boldsymbol{\Phi}^T\Big|^{-\frac{1}{2}} \\ &\quad \cdot\exp\left(-\frac{1}{2}\boldsymbol{\Delta X}^T\Big(\sigma^2\tau\boldsymbol{I} + \boldsymbol{\Phi A}^{-1}\boldsymbol{\Phi}^T\Big)^{-1}\boldsymbol{\Delta X}\right).\end{aligned} \tag{19}$$

In fact, it is difficult to acquire the optimal values of $\boldsymbol{\alpha}$ and σ maximizing (19) in the closed form, but in this case, the equation can be derived to iteratively re-estimate the solution of the optimization problem.

$\boldsymbol{\alpha}$ is equal with differentiation of (19) and with 0. According to the method in [26], the following can be obtained:

$$\alpha_i^{\mathrm{new}} = \frac{\gamma_i}{\mu_i^2} \tag{20}$$

where μ_i is the ith posterior mean weight of $\boldsymbol{\mu}$, and γ_i is defined by

$$\gamma_i = 1 - \alpha_i\Sigma_{ii} \tag{21}$$

where Σ_{ii} refers to the ith diagonal element of the covariance of posterior weight from (17) calculated with the current $\boldsymbol{\alpha}$ and σ^2. In order to update σ^2, the differentiation increases the re-estimate

$$\Big(\sigma^2\tau\Big)^{\mathrm{new}} = \frac{\|\boldsymbol{\Delta X} - \boldsymbol{\Phi\mu}\|^2}{m - \Sigma_{i=1}^m \gamma_i} \tag{22}$$

where m refers to the amount of data samples.

Therefore, if (20) and (21) are applied repeatedly, the algorithm can simultaneously update the posterior statistics μ and Σ from (16) and (17) until the appropriate convergence criterion is satisfied. As a result, the mean weights $\boldsymbol{\omega}$ by (16) and σ by (22) can be estimated. In other words, both $\boldsymbol{\theta_{\mathrm{RVM}}}$ and σ in the degradation model have been calculated.

When the parameter estimation procedure converges, given a new test point t_*, the prediction on basis of the posterior distribution over ω, the conditioned on the maximizing values μ_{MP} and σ_{MP}^2 can be found for the corresponding degradation increment target ΔX_*. Concretely, the distribution of ΔX_* is a normal distribution like

$$p\Big(\Delta X_*|\Delta X, \boldsymbol{\alpha}_{\mathbf{MP}}, \sigma_{\mathrm{MP}}^2\tau\Big) = N\Big(\Delta X_*|f_*, \sigma_*^2\tau\Big) \tag{23}$$

with the mean f_* and the variance $\sigma_*^2\tau$ with

$$f_* = \boldsymbol{\mu}^T\boldsymbol{\phi}(\boldsymbol{t_*}) \tag{24}$$

$$\sigma_*^2\tau = \sigma_{\mathrm{MP}}^2\tau + \boldsymbol{\phi}(t_*)^T\boldsymbol{\Sigma}\boldsymbol{\phi}(t_*). \tag{25}$$

The predictive mean is the sum of basis functions, weighted by the posterior mean weights, where the majority of functions will generally become 0 to achieve the sparsity. The training vectors corresponding to the remaining nonzero weights are referred as the relevance vectors.

Thanks to the sparse Bayesian learning algorithm of the RVM model, $\boldsymbol{\theta}_{\mathbf{RVM}}$ and σ have been calculated by using the probabilistic Bayesian learning framework to set the foundation for the following RUL estimation.

IV. RUL Estimation

Based on the aforementioned "intelligent" degradation process model and the calculated parameters, the derivation of the RUL distribution is investigated. First, it is necessary to introduce the definition of lifetime and RUL. The lifetime is generally defined as the FHT of the degradation process crossing a preset threshold level ξ

$$T = \inf\{t : X(t) \geq \xi | X(0) \leq \xi\}. \tag{26}$$

Under the concept of the FHT, the RUL at t_k is defined as

$$L_k = \inf\{l_k : X(t_k + l_k) \geq \xi | x(t_k) \leq \xi\}. \tag{27}$$

Considering the stochastic dynamics in the degradation process, both the lifetime T and RUL L_k defined by (26) and (27) are random variables. The PDFs of T and the RUL L_k provide essential instructions for decision making in PHM. In order to obtain PDF calculation formulas of T and L_k, a lemma in [13] is presented first.

Lemma 1: The PDF of the FHT for $X(t)$ crossing a constant boundary can be approximated with an explicit form as

$$p_{X(t)}(\xi, t) \cong \frac{1}{\sqrt{2\pi t}} \cdot \left(\frac{S(t)}{t} + \frac{1}{\sigma} \cdot \mu(t; \boldsymbol{\theta})\right) \cdot \exp\left(-\frac{S^2(t)}{2t}\right) \tag{28}$$

where the time-varying boundary is

$$S(t) = \frac{1}{\sigma} \cdot \left(\xi - \int_0^t \mu(\tau; \boldsymbol{\theta}) d\tau\right). \tag{29}$$

Before Theorems 1 and 2 are derived, the following important assumption needs to be made.

Assumption 1: The CM point t_k, FHT T, and RUL L_k satisfy $t_k = k_t\tau, k_t \in N^+$, $T = k_T\tau, k_T \in N^+$, and $L_k = k_L\tau, k_L \in N^+$, respectively, where τ is the step length of the degradation increment.

Remark 1: Assumption 1 takes into account the discretization of the infinite, continuous space for the numerical approximation calculation of RUL distribution. Most of the RUL estimation methods are based on the sample data from sensors, monitoring the degradation process, and the estimation accuracy is verified by comparing the estimated RUL with the real RUL from the observation samples. The proposed method sets, in general, the integer multiples of the sample rate as the step length τ. τ is one of the hyperparameters in our model, it should be selected by cross validation.

Using Lemma 1 and based on Assumption 1, the following approximation of the FHT distribution can be derived at the CM point t_k.

Theorem 1: Set $g(k) = \sum_{j=0}^{k-1}\int_{j\tau}^{(j+1)\tau}\mu(t; \boldsymbol{\theta})dt$, under Assumption 1, and the PDF value of the FHT of degradation state $X(t)$ crossing a constant threshold ξ can be approximated numerically as

$$\begin{aligned} f_T(k_T\tau) = {}& \frac{1}{\sqrt{2\pi k_T\tau}} \cdot \exp\left(-\frac{S^2(k_T\tau)}{2k_T\tau}\right) \\ & \cdot \left(\frac{S(k_T\tau)}{k_T\tau} + \frac{1}{\sigma} \cdot \frac{g(k_T+1) - g(k_T-1)}{2\tau}\right) \end{aligned} \tag{30}$$

where

$$S(k_T\tau) = \frac{1}{\sigma} \cdot (\xi - g(k_T)). \tag{31}$$

Proof: Based on Assumption 1, if the PDF of T for $X(t)$ is

$$\begin{aligned} p_{X(T)}(\xi, T) \cong {}& \frac{1}{\sqrt{2\pi T}} \cdot \left(\frac{S(T)}{T} + \frac{1}{\sigma}\mu(T; \boldsymbol{\theta})\right) \\ & \cdot \exp\left(-\frac{S^2(T)}{2T}\right) \end{aligned} \tag{32}$$

the central difference to estimate $\mu(T; \boldsymbol{\theta})$ can be applied

$$\begin{aligned} \mu(T; \boldsymbol{\theta}) &= \frac{d\int_0^T \mu(t; \boldsymbol{\theta})dt}{dT} \\ &\cong \frac{\int_0^{T+\tau}\mu(t; \boldsymbol{\theta})dt - \int_0^{T-\tau}\mu(t; \boldsymbol{\theta})dt}{2\tau} \\ &= \frac{\int_0^{(k_T+1)\tau}\mu(t; \boldsymbol{\theta})dt - \int_0^{(k_T-1)\tau}\mu(t; \boldsymbol{\theta})dt}{2\tau} \\ &= \frac{\sum_{j=0}^{k_T}\int_{j\tau}^{(j+1)\tau}\mu(t; \boldsymbol{\theta})dt - \sum_{j=0}^{k_T-2}\int_{j\tau}^{(j+1)\tau}\mu(t; \boldsymbol{\theta})dt}{2\tau} \\ &= \frac{g(k_T+1) - g(k_T-1)}{2\tau}. \end{aligned} \tag{33}$$

The time-varying boundary $S(T)$ is then represented numerically as

$$\begin{aligned} S(T) &= \frac{1}{\sigma} \cdot \left(\xi - \int_0^T \mu(t; \boldsymbol{\theta})dt\right) \\ &= \frac{1}{\sigma} \cdot (\xi - g(k_T)). \end{aligned} \tag{34}$$

This completes the proof of Theorem 1. ■

Then, the main application of the proposed modeling method will be introduced, which is the RUL estimation at a specific CM point in observation time. So, another theorem is proposed as follows.

Theorem 2: Set $g(k) = \sum_{j=0}^{k-1}\int_{j\tau}^{(j+1)\tau}\mu(t; \boldsymbol{\theta})dt$, under Assumption 1, given the available degradation measurement $X(t_k)$, the PDF value of the RUL of $X(t)$ can be formulated numerically at time t_k as

$$\begin{aligned} f_L(k_L\tau) = {}& \frac{1}{\sqrt{2\pi k_L\tau}} \cdot \exp\left(-\frac{S^2(k_L\tau)}{2k_L\tau}\right) \\ & \cdot \left(\frac{S(k_L\tau)}{k_L\tau} + \frac{1}{\sigma} \cdot \frac{g(k_L+1) - g(k_L-1)}{2\tau}\right) \end{aligned} \tag{35}$$

Algorithm 1 RUL Estimation

Input:
$X_{t_1:t_k} = \{X(t_1), \dots, X(t_k)\}$: History of the degradation measurements to t_k;

Output:
$F_L(t_k) = \{f_L(k_L\tau), k_L = 1, 2, \dots\}$: PDF values of the RUL at t_k;

Determine the step length of degradation increment τ and the basis function form, then obtain the degradation increment data set $\{t_k, \Delta X_\tau(t_k)\}_{k=1}^m$;
Initialize $\{\alpha_i\}$ and σ^2;
repeat
  Compute weight posterior statistics $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by (16) and (17);
  Compute all $\{\gamma_i\}$ by (21), then re-estimate $\{\alpha_i\}$ and σ^2 by (20) and (22);
until convergence
make prediction for new drift increment by (23);
Delete basis functions for an optimal $\alpha_i = \infty$;
compute PDF values of the RUL at t_k: $F_L(t_k) = \{f_L(k_L\tau), k_L = 1, 2, ...\}$ by (35);
return $F_L(t_k)$;

where

$$S(k_L\tau) = \frac{1}{\sigma} \cdot (\xi - X(k_t\tau) - (g((k_t + k_L)\tau) - g(k_t\tau))). \quad (36)$$

Proof: Similar to the proof in [13], as the measurement $X(t_k)$ has been observed, for $t \geq t_k$, the degradation process is rewritten as $X(t) = X(t_k) + \int_{t_k}^{t} \mu(t; \boldsymbol{\theta})dt + \sigma B(t - t_k)$. The transformation $l_k = t - t_k$ with $l_k \geq 0$ is applied, and then the process $\{X(t), t \geq t_k\}$ changes into $X(l_k + t_k) - X(t_k) = \int_{t_k}^{l_k+t_k} \mu(t; \boldsymbol{\theta})dt + \sigma B(l_k)$, with $l_k \geq 0$.

Consequently, the RUL at time t_k equals the FHT of the process $\{Y(l_k), l_k \geq 0\}$, crossing the threshold $\xi_{t_k} = \xi - X(t_k)$, where $Y(l_k) = X(l_k + t_k) - X(t_k)$ and $Y(0) = 0$.

This leads to

$$Y(l_k) = \int_{t_k}^{l_k+t_k} \mu(t; \boldsymbol{\theta})dt + \sigma B(l_k). \quad (37)$$

Using Lemma 1, Assumption 1, and Theorem 1 results in

$$\mu(L_k; \boldsymbol{\theta}) = \frac{g(k_L + 1) - g(k_L - 1)}{2\tau} \quad (38)$$

where $g(k) = \sum_{j=0}^{k-1} \int_{j\tau}^{(j+1)\tau} \mu(t; \boldsymbol{\theta})dt$, and

$$S(k_L\tau) = \frac{1}{\sigma}(\xi - X(k_t\tau) - (g((k_t + k_L)\tau) - g(k_t\tau))) \quad (39)$$

which completes the proof of Theorem 2. ■

The RUL estimation algorithm is summarized as Algorithm 1.

V. Experiments

In this section, the application of the introduced method for the RUL estimation is demonstrated by two examples: 1) a numerical simulation and 2) a practical degraded bearing case study. In addition, the proposed method is compared with the existing state-of-the-art ones, i.e., the power and exponential law model in [5] and [13]. All the experiments are performed using MATLAB (R2014b) on a computer with an Intel Core i5-4460 CPU with 3.20 GHz and 8.00-GB RAM.

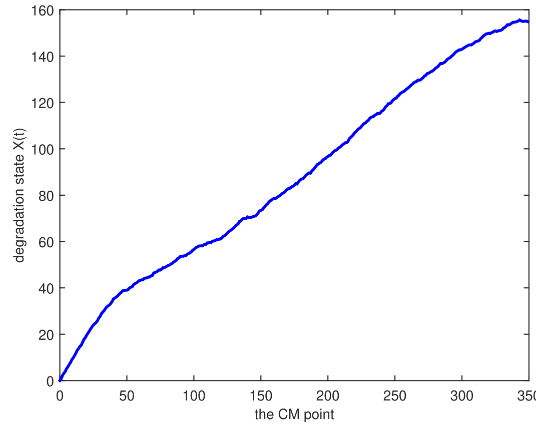


Fig. 2. Degradation data of D1.

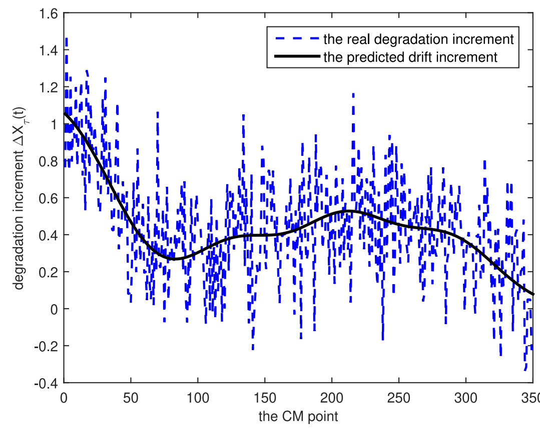


Fig. 3. Real degradation increment and the predicted drift increment (the predictive mean of degradation increment) with RVM model using D1.

A. Numerical Simulation

Three sets of numerically simulated data are utilized to demonstrate the validity of the proposed method in comparison with the power-law model (Si-1) and the exponential model (Si-2) presented in [13].

1) *D1:* The degradation increments are driven by $\Delta X_\tau(t) = \int_t^{t+\tau} \mu(t; \boldsymbol{\theta})dt + \eta(t)$ where the drift increment is based on a combination of Gaussian basis functions $\exp(-([(t - t_i)^2]/q^2))$ with $q = 50$ and the weights are made up of most zeros and a small amount of normally distributed pseudorandom numbers. As shown in Fig. 2, the degradation state $X(t)$ is assumed to reach the preset failure threshold at the 350th CM point along such a special nonlinear stochastic degradation trajectory.
2) *D2:* The simulated degradation data are generated by (2) following Si-1 model where $a = 1.5, b = 0.7$, and $\sigma = 2.7$. The degradation state $X(t)$ is assumed to reach the failure threshold at the 500th CM point.
3) *D3:* The simulated degradation data is generated by (3) following Si-2 model where $a = 1.5, b = 0.01$, and

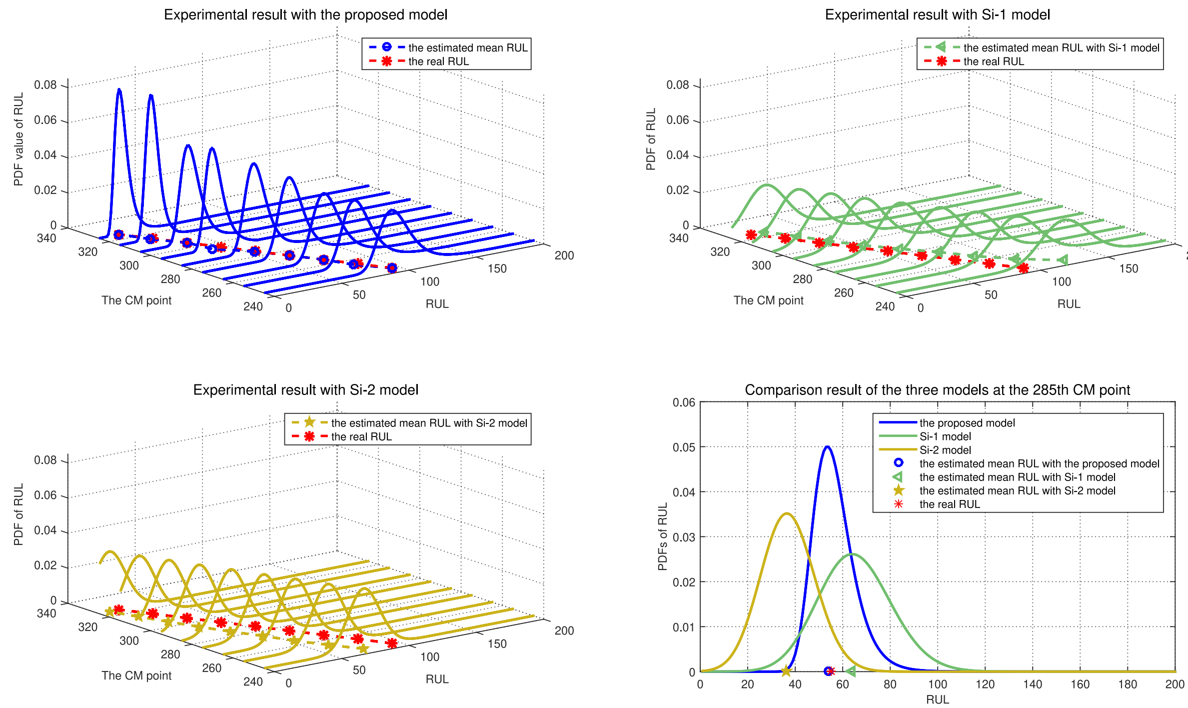


Fig. 4. Comparison of the RUL estimation results based on the proposed model, the Si-1 and Si-2 models with D1.

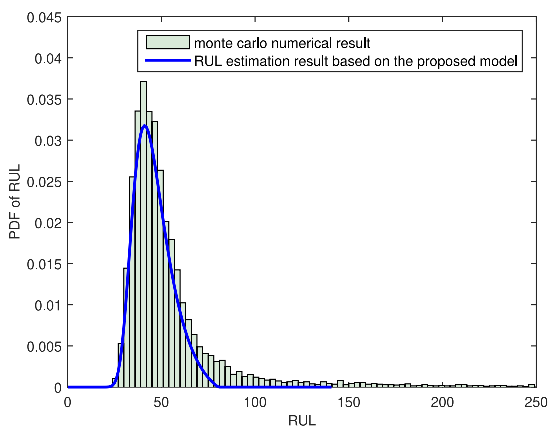


Fig. 5. Comparison of the PDF of the proposed model and Monte Carlo numerical result at the 300th CM point with D1.

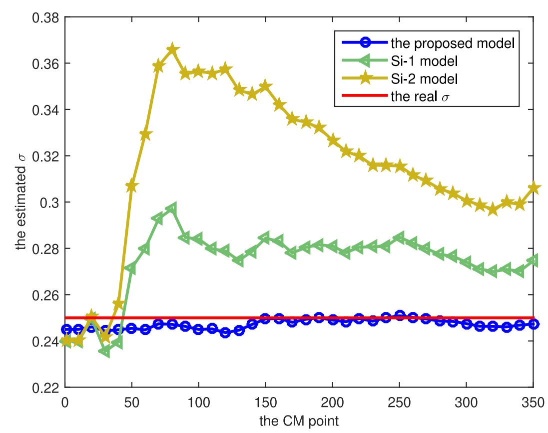


Fig. 6. Comparison of the estimation results of σ based on the proposed model, the Si-1 and Si-2 models from 1st to 350th CM point with D1.

$\sigma = 2.7$. $X(t)$ is assumed to reach the failure threshold at the 500th CM point.

First, D1 is used to compare the performance of the proposed model and Si-1 and Si-2 models. Preparing for modeling the degradation process as (5), some basic data preprocessing work is applied to obtain the degradation increment data with $\tau = 1$ shown by the blue dotted line in Fig. 3. Besides, the black solid line represents the drift increment predicted by the sparse Bayesian learning algorithm with the degradation increment data, which shows that the prediction result well fits the real nonlinear change trend of degradation increment. While the degradation increment is predicted, the estimation of diffusion coefficient $\hat{\sigma}$ is completed by the sparse Bayesian learning algorithm that $\hat{\sigma}^{-2} = 16.3$ at the 300th CM point.

Algorithm 1 is suitable for the online RUL estimation of degradation data. The RUL estimation result at the ith CM point is derived by the data of the first i CM points. In the numerical simulation, Algorithm 1 is used to conduct the RUL estimation every ten points from the 245th to 325th CM point. To be specific, the data of the first 245 CM points (70% of the complete degradation data) are utilized to model the degradation and estimate the RUL online at the 245th CM points. The average computing time is 0.085 s, approximately. As more CM degradation data are available, the RUL estimation result is updated based on new online data by Algorithm 1. The comparison among results based on the proposed model, the power-law model (Si-1) and the exponential model (Si-2) presented in [13] are showed in Fig. 4. MLE is applied to compute the unknown parameters in Si-1 and Si-2 by the MATLAB function "fminsearch." the RUL estimation results based on the Si-1 and Si-2 model according to the theorem in [13] have much more biases from the real RUL (see Fig. 4). In order to compare the superiority and inferiority of the three

TABLE I
PARAMETER CALCULATION RESULTS AT 300TH CM POINT USING D1, D2, AND D3

dataset	model	a	b	$\boldsymbol{\omega}$	σ
D1	the proposed model	-	-	$\omega_{16} = 0.6148; \omega_{127} = 0.3515; \omega_{210} = 0.4727; \omega_{288} = 0.3648$	0.2475
	Si-1 model	1.7339	0.7673	-	0.2740
	Si-2 model	2.7956e7	1.7015e-8	-	0.3007
D2	the proposed model	-	-	$\omega_1 = -1.5736; \omega_2 = 1.6148$	2.5837
	Si-1 model	1.3502	0.8459	-	2.7509
D3	the proposed model	-	-	$\omega_1 = 0.0214$	2.5803
	Si-2 model	1.7855	0.0058	-	2.5832

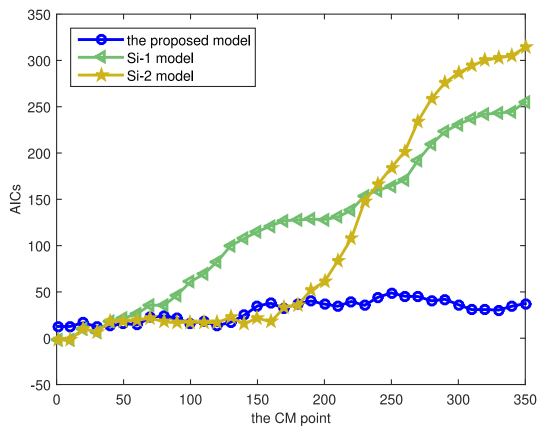


Fig. 7. Comparison of AICs of the proposed model, the Si-1 and Si-2 model from 1st to 350th CM point with D1.

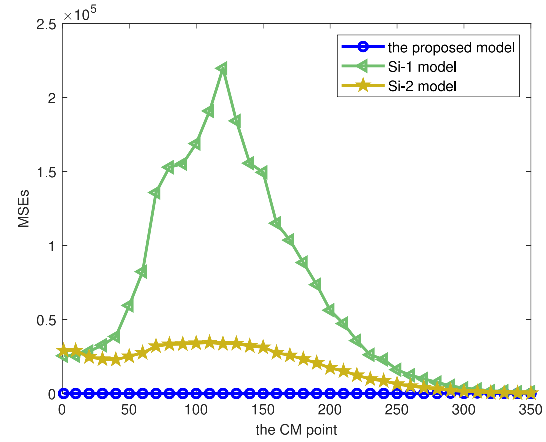


Fig. 8. Comparison of MSEs of RUL estimation results based on the proposed model, Si-1 and Si-2 models from 1st to 350th CM point with D1.

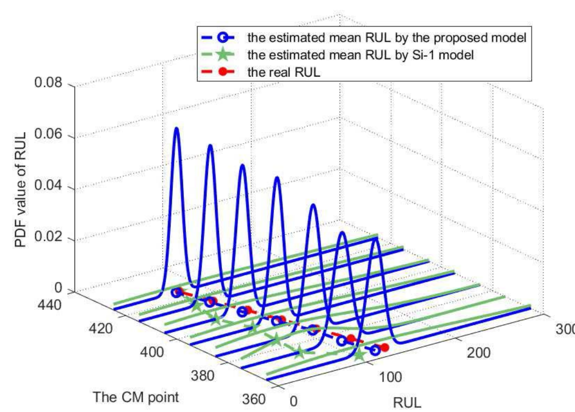


Fig. 9. Comparison of the RUL estimation results based on the proposed model and Si-1 model with D2.

models more intuitively, the PDFs of the RUL based on the three models at the 285th CM time point is plotted in the same figure (see Fig. 4). Compared to the Si-1 and Si-2 models, the estimated PDF of the RUL under the proposed model has less uncertainty, as shown in Fig. 4. In addition, the location of the RUL distribution is closer to the actual RUL.

Moreover, the proposed modeling method is compared with the Monte Carlo simulation result to verify its effectiveness. The estimation results are compared to those calculated by the Monte Carlo method with 10 000 simulated sample paths (see Fig. 5), which indicates that the RUL estimation result is basically consistent with the numerical RUL distribution as well as the proposed method can approximate the PDF of the RUL well.

In addition, the proposed model, the Si-1 and Si-2 models are compared from different points of view, including the parameter estimation, the fitting and the RUL estimation error. In terms of parameter estimation, Compared with the Si-1 and Si-2 models, the diffusion coefficience σ estimated by the sparse Bayesian learning algorithm of the proposed model at each CM point is closer to the true value 0.25 (see Fig. 6). The remaining parameters of the Si-1 and Si-2 models calculated at the 300th CM point are also listed in the Table I.

In order to compare the fitting of the proposed model, the Si-1 and Si-2 model, the Akaike information criterion (AIC) [31] of the fitted model is used for evaluating the fitting goodness and complexity of each model. The AIC at t_k is formulated by

$$\mathrm{AIC}_k = -2(\max l_k) + 2P_k \tag{40}$$

where P_k refers to the quantity of predicted model parameters at t_k, and $\max l_k$ represents the maximized likelihood at t_k. According to the definition, a smaller AIC means a better fitting performance. The comparison of AICs of these three models at each CM point in Fig. 7 illustrates the better fitting ability of the proposed model.

On the part of RUL estimation error, the mean-squared error (MSE) is employed to assess performance of the proposed model, the Si-1 and Si-2 model. The MSE at t_k is calculated as

$$\mathrm{MSE}_k = \int_0^\infty \left(l_k - \tilde{l}_k\right)^2 f_L(l_k) dl_k \tag{41}$$

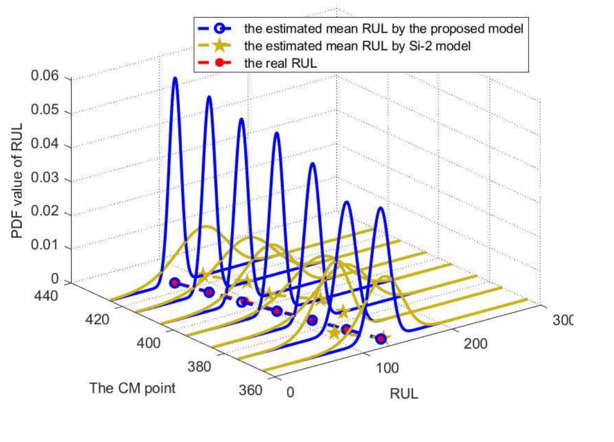


Fig. 10. Comparison of the RUL estimation results based on the proposed model and Si-2 model with D3.

where $\tilde{l}_k = T_\xi - t_k$ is the bona fide RUL at t_k, and $f_L(\cdot)$ is the PDF of the RUL. As revealed in Fig. 8, the proposed model develops a more suitable RUL estimation with smaller MSE at each CM point compared with the Si-1 and Si-2 models.

Second, D2 is used to compare the performance of the proposed model and Si-1 model. To establish the proposed model, set $\tau = 1$, and select the polynomial basis function where $p = 0.7$. Table I shows the results of parameter calculation using D2. As shown in Fig. 9, the comparison of the RUL estimation results every ten points from the 365th to 425th CM point illustrates that the performance of the proposed model for RUL estimation is better.

Third, D3 is used to compare the performance of the proposed model and Si-2 model. Following the proposed model, τ is set as 1, and the exponential basis function is selected with $q = 100$. Table I shows the parameter calculation results using D3. The comparison of the RUL estimation results in Fig. 10 shows the proposed model provides more accurate RUL estimation with less uncertainty than the Si-2 model.

Therefore, the superiority of the proposed method in stochastic nonlinear degradation process modeling and RUL estimation has been demonstrated. This method can be a better aid to PHM when some systems only have or also have this type of special nonlinear degradation measurement signals.

B. Practical Case Study

In this part, a practical case is used to investigate the RUL estimation performance of the proposed model. The bearing degradation dataset [27] is utilized to compare the performance of the RUL estimation based on the proposed model and the model in [5] (denoted as the Huang model). Taking the bearing1_1 for example, the vibration and the temperature signals of the bearing were gathered under the operating condition of 1800 rpm and 4000 N on a laboratory experimental platform (PRONOSTIA). The sampling frequency of the vibration and temperature signal is 25.6 kHz and 10 Hz, respectively. 2560 consecutive samples of the vibration sensor are recorded every 10 s. Since the sampling interval is 10 s, here, the degradation increment step length τ is set as 10 s. This makes the CM

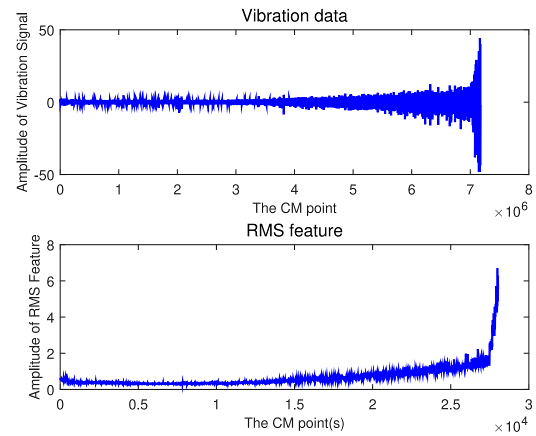


Fig. 11. Vibration signal and RMS feature data of bearing1_1.

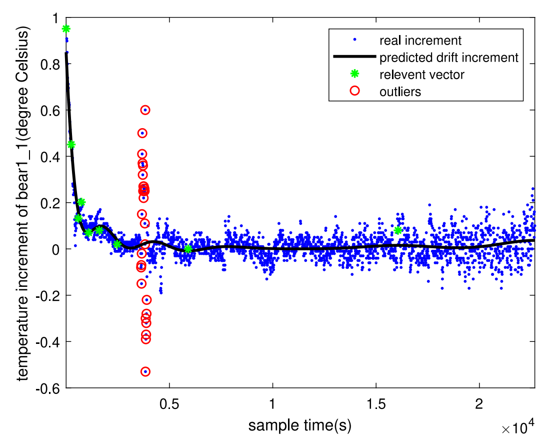


Fig. 12. Bearing1_1 temperature increment data.

point t_k, the FHT T, and the RUL L_k can share one degradation increment step length. In order to estimate the RUL of the bearing1_1, similar to [5], the root mean square (RMS) feature for the recorded vibration data is extracted first, after which a new time series of degradation data is obtained.

According to the definition of failure in [5], once the amplitude of a raw vibration signal overpasses 20 g, the bearing under test is identified as reaching the failure threshold. For the bearing1_1, the raw vibration signals overpass 20 g at its 2756th sampling set. Therefore, the corresponding RMS value and temperature measurement are 2.6762 and 156.51 °C, respectively, which are defined as the failure threshold of the bearing degradation process in the RMS feature domain and the temperature measurement domain. The completed life duration of the bearing1_1 is considered as 27 560 s, when the degradation trajectory is truncated after the sample at 27 560 s.

The temperature signal, as well as the raw horizontal vibration signal and its corresponding RMS feature, are plotted in Figs. 1 and 11, respectively. Compared with the RMS feature of the vibration signal (see Fig. 11), the temperature signal (see Fig. 1) depicts a degradation trajectory with more nonlinear behavior. Huang *et al.* [5] established a nonlinear heterogeneous Wiener process model with an adaptive drift (Huang model) using the RMS feature data by specifying a time-transformation function with the form of power law $\Lambda(t;\xi) = t^b$. The model in [5] can be used to deal with the nonlinearity of the RMS feature of vibration signals, but it may

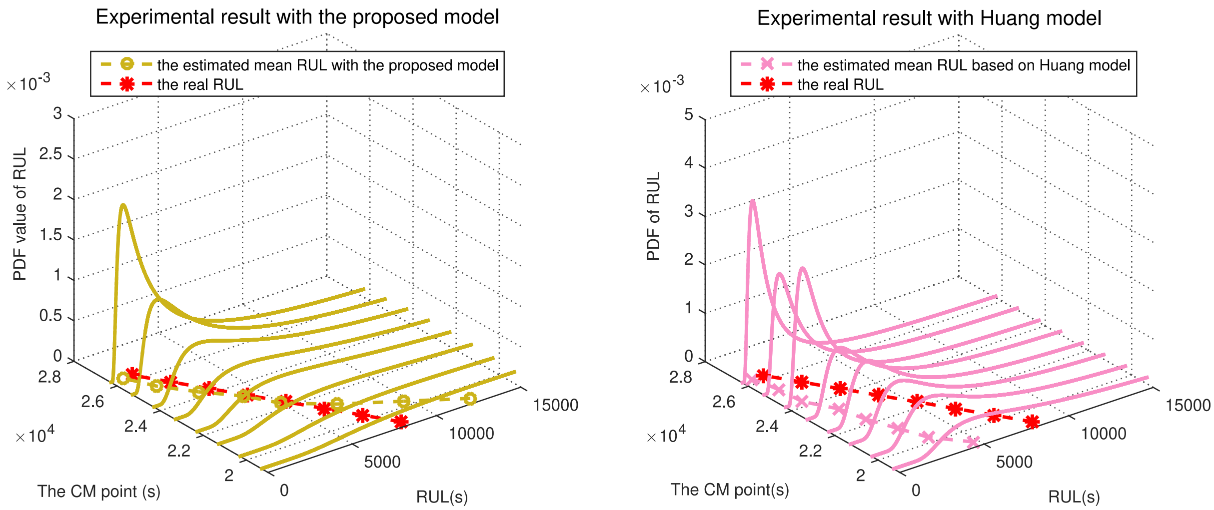


Fig. 13. Comparison of the RUL estimation results of the bearing1_1 based on the proposed model and Huang model from CM point 19 300 to 26 300 s using vibration data.

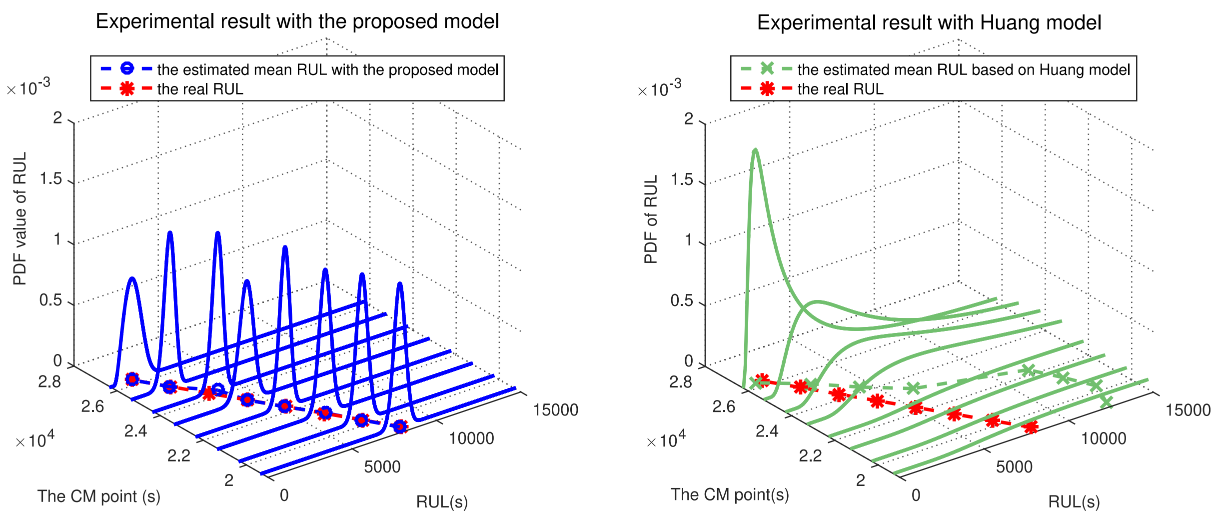


Fig. 14. Comparison of the RUL estimation results of the bearing1_1 based on the proposed model and Huang model from CM point 19 300 to 26 300 s using temperature data.

not be suitable for some special Wiener processes with more nonlinear behavior, such as the temperature signal in Fig. 1.

When the vibration and temperature data are processed to estimate RUL using Algorithm 1, the number of iterations does not exceed 200 at most. The longest consumed time for the vibration and temperature data is 0.09 and 0.46 s, respectively. Both are smaller than the step length of degradation increment τ (10 s) for processing the vibration and temperature data, which illustrates that the proposed algorithm is suitable for the real-time online application.

In the application of the proposed method, the sparse Bayesian algorithm can avoid the influence of outliers on the degradation modeling and ensure the robustness of the method. In Fig. 12, the samples from 3640 to 3900 s circled in red deviate significantly from the trend of degradation increment based on the temperature signal and are considered as outliers. The sparse Bayesian algorithm is used to model the degradation increment. Then, nine samples marked as * are selected as relevant vectors to establish the degradation model and predict the drift increment as shown in Fig. 12. The outliers are not included in the relevant vectors. Their corresponding weights are zeros. Therefore, the outliers do not affect the degradation modeling, and the robustness of the proposed method is demonstrated.

In the practical case study, the RUL estimation is conducted every 1000 s from CM point 19 300 to 26 300 s based on the proposed model and Huang model [5], which means 70%–95% of the complete degraded bearing vibration and temperature measurement data is used to train the models and estimate the RULs of bearing. The Gaussian and polynomial basis function are used to learn the proposed models with the temperature and vibration signal data, respectively. Figs. 13 and 14 show the RUL estimation results based on the proposed model and Huang model using vibration and temperature measurement data, respectively. Both of them illustrate that the proposed model highlights the advantages compared with the Huang model, whether using vibration data or temperature data. The advantages include a more accurate RUL estimation and a less uncertainty of estimation.

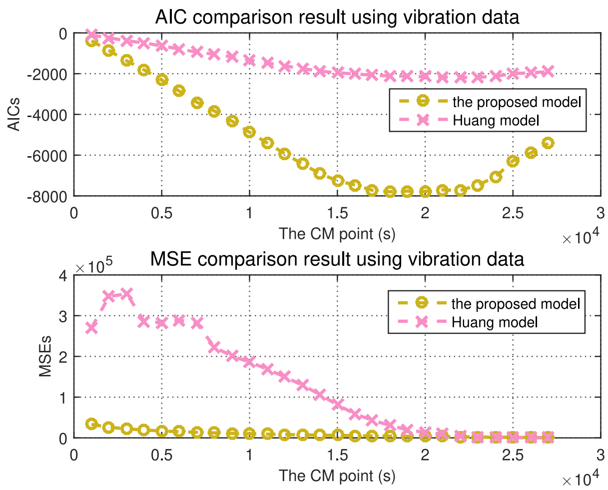


Fig. 15. Comparison of AIC and MSE between the proposed model and Huang model using vibration data of bearing1_1.

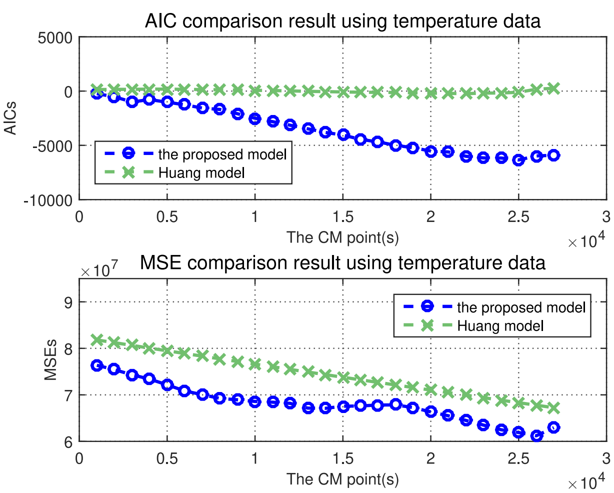


Fig. 16. Comparison of AIC and MSE between the proposed model and Huang model using temperature data of bearing1_1.

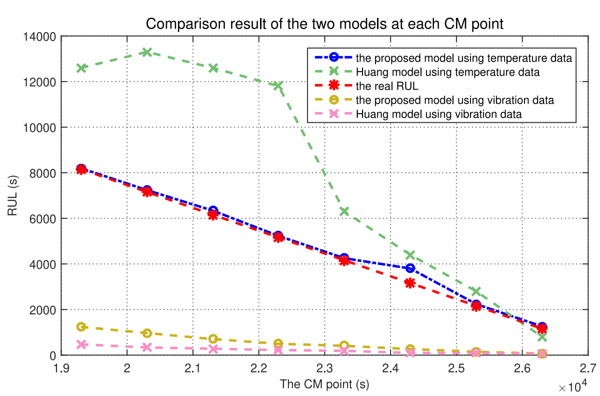


Fig. 17. Comparison of the RUL estimation results of the bearing1_1 based on the proposed model and Huang model from CM point 19 300 to 26 300 s using temperature and vibration data.

To indicate the superirity of the proposed model, AICs and MSEs calculation comparison results using vibration data or temperature data between the proposed model and Huang model are illustrated in Figs. 15 and 16, respectively, where smaller AICs and MSEs at each CM point demonstrate the proposed model's better fitting ability and superior RUL estimation performance.

In addition, all the estimated mean RULs at each CM point are compared in Fig. 17. It can be seen that the proposed method does not show obvious advantage over Huang's model when dealing with vibration data, while it significantly improves the estimation accuracy when temperature data are adopted. Based on these observation, the method proposed is more qualified to model degradation processes, which include more nonlinear behavior.

VI. Conclusion

This article proposed a new and data-driven modeling and RUL estimation method for the stochastic degradation processes with more complicated nonlinear behavior. An RVM-based increment was utilized to depict the nonlinearity of degradation process which basic modeling methods cannot handle well. The sparse Bayesian learning algorithm was adopted to compute the unknown parameters in the degradation process. Then, an algorithm for RUL estimation was presented. A numerical simulation and a practical case of a degraded bearing demonstrate the validity of the proposed method and the accuracy of RUL estimation. The simulation results showed the proposed modeling method provides more reliable performance and accuracy RUL estimation result than the conventional nonlinear degradation process modeling methods. Although the proposed modeling method has greatly liberated the constraint of nonlinear expression, it is still a parametric modeling method, which needs to preset the basis function form. A further improvement could be achieved by some nonparametric modeling methods for the drift coefficient $\mu(t;\boldsymbol{\theta})$ in the future in order to enrich the applications of nonlinear degradation models based on the Wiener process. Besides, in practical engineering, it is a more practical case that the degradation variables are not fully accessible with the improvement of the equipment complexity. Therefore, another future work is to consider the RUL estimation when the degradation is not completely known.

References

[1] M. Pecht, "Prognostics and health management of electronics," in *Encyclopedia of Structural Health Monitoring*. Chichester, U.K.: Wiley, 2009.

[2] 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, 2018.

[3] F. Tamssaouet, K. T. P. Nguyen, and K. Medjaher, "System-level prognostics under mission profile effects using inoperability input-output model," *IEEE Trans. Syst., Man, Cybern., Syst.*, early access, Oct. 16, 2019, doi: 10.1109/TSMC.2019.2944834.

[4] Y. Zhang, R. Xiong, H. He, and M. G. Pecht, "Lithium-ion battery remaining useful life prediction with box–cox transformation and Monte Carlo simulation," *IEEE Trans. Ind. Electron.*, vol. 66, no. 2, pp. 1585–1597, Feb. 2019.

[5] Z. Huang, Z. Xu, W. Wang, and Y. Sun, "Remaining useful life prediction for a nonlinear heterogeneous Wiener process model with an adaptive drift," *IEEE Trans. Rel.*, vol. 64, no. 2, pp. 687–700, Jun. 2015.

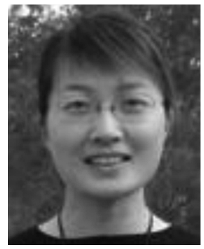
[6] J. Deutsch and D. He, "Using deep learning-based approach to predict remaining useful life of rotating components," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 1, pp. 11–20, Jan. 2018.

[7] C.-H. Hu, H. Pei, X.-S. Si, D.-B. Du, Z.-N. Pang, and X. Wang, "A prognostic model based on DBN and diffusion process for degrading bearing," *IEEE Trans. Ind. Electron.*, vol. 67, no. 10, pp. 8767–8777, Oct. 2020.

[8] J. I. Aizpurua, S. D. J. McArthur, B. G. Stewart, B. Lambert, J. G. Cross, and V. M. Catterson, "Adaptive power transformer lifetime predictions through machine learning and uncertainty modeling in nuclear power plants," *IEEE Trans. Ind. Electron.*, vol. 66, no. 6, pp. 4726–4737, Jun. 2019.
[9] J. I. Aizpurua, V. M. Catterson, I. F. Abdulhadi, and M. S. Garcia, "A model-based hybrid approach for circuit breaker prognostics encompassing dynamic reliability and uncertainty," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 9, pp. 1637–1648, Sep. 2018.
[10] M. J. Mcghee, V. M. Catterson, and B. Brown, "Prognostic modeling utilizing a high-fidelity pressurized water reactor simulator," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 5, pp. 727–732, May 2018.
[11] N. Li, Y. Lei, T. Yan, N. Li, and T. Han, "A Wiener-process-model-based method for remaining useful life prediction considering unit-to-unit variability," *IEEE Trans. Ind. Electron.*, vol. 66, no. 3, pp. 2092–2101, Mar. 2019.
[12] Q. Guan, Y. Tang, and A. Xu, "Objective Bayesian analysis accelerated degradation test based on Wiener process models," *Appl. Math. Model.*, vol. 40, no. 4, pp. 2743–2755, 2016.
[13] X.-S. Si, W. Wang, C.-H. Hu, D.-H. Zhou, and M. G. Pecht, "Remaining useful life estimation based on a nonlinear diffusion degradation process," *IEEE Trans. Rel.*, vol. 61, no. 1, pp. 50–67, Mar. 2012.
[14] J. H. Cha and G. Pulcini, "Optimal burn-in procedure for mixed populations based on the device degradation process history," *Eur. J. Oper. Res.*, vol. 251, no. 3, pp. 988–998, 2016.
[15] M. H. Ling, H. K. T. Ng, and K. L. Tsui, "Bayesian and likelihood inferences on remaining useful life in two-phase degradation models under gamma process," *Rel. Eng. Syst. Safety*, vol. 184, pp. 77–85, Apr. 2019.
[16] Z.-S. Ye and N. Chen, "The inverse Gaussian process as a degradation model," *Technometrics*, vol. 56, no. 3, pp. 302–311, 2014.
[17] W. Peng, Y.-F. Li, Y.-J. Yang, H.-Z. Huang, and M. J. Zuo, "Inverse Gaussian process models for degradation analysis: A Bayesian perspective," *Rel. Eng. Syst. Safety*, vol. 130, pp. 175–189, Oct. 2014.
[18] M. Wei, M. Chen, and D. Zhou, "Multi-sensor information based remaining useful life prediction with anticipated performance," *IEEE Trans. Rel.*, vol. 62, no. 1, pp. 183–198, Mar. 2013.
[19] Z. Hanwen, C. Maoyin, and Z. Donghua, "Remaining useful life prediction for a nonlinear multi-degradation system with public noise," *J. Syst. Eng. Electron.*, vol. 29, no. 2, pp. 429–435, Apr. 2018.
[20] Q. Zhai and Z.-S. Ye, "RUL prediction of deteriorating products using an adaptive Wiener process model," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 2911–2921, Dec. 2017.
[21] 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, 2005.
[22] G. Whitmore, "Estimating degradation by a Wiener diffusion process subject to measurement error," *Lifetime Data Anal.*, vol. 1, no. 3, pp. 307–319, 1995.
[23] K. Le Son, M. Fouladirad, A. Barros, E. Levrat, and B. Iung, "Remaining useful life estimation based on stochastic deterioration models: A comparative study," *Rel. Eng. Syst. Safety*, vol. 112, pp. 165–175, Apr. 2013.
[24] Z. Zhang, C. Hu, X. Si, J. Zhang, and J. Zheng, "Stochastic degradation process modeling and remaining useful life estimation with flexible random-effects," *J. Frankl. Inst.*, vol. 354, no. 6, pp. 2477–2499, 2017.
[25] Y. Deng, A. Barros, and A. Grall, "Degradation modeling based on a time-dependent ornstein-uhlenbeck process and residual useful lifetime estimation," *IEEE Trans. Rel.*, vol. 65, no. 1, pp. 126–140, Mar. 2016.
[26] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Jun. 2001.
[27] P. Nectoux *et al.*, "PRONOSTIA: An experimental platform for bearings accelerated degradation tests," in *Proc. IEEE Int. Conf. Prognostics Health Manag. (PHM)*, 2012, pp. 1–8.
[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] J. Tang and T.-S. Su, "Estimating failure time distribution and its parameters based on intermediate data from a Wiener degradation model," *Naval Res. Logist.*, vol. 55, no. 3, pp. 265–276, 2008.
[30] I. Karatzas and S. E. Shreve, "Brownian motion," in *Brownian Motion and Stochastic Calculus*. New York, NY, USA: Springer, 1998, pp. 47–127.
[31] H. Akaike, "A new look at the statistical model identification," in *Selected Papers of Hirotugu Akaike*. New York, NY, USA: Springer, 1974, pp. 215–222.

Yuhan Zhang received the B.E. degree in automation and electronics engineering from the University of Science and Technology Beijing, Beijing, China, in 2017. She is currently pursuing the Ph.D. degree in dynamical systems and control with Peking University, Beijing.

Her current research interests focus on data-driven fault diagnosis and PHM.

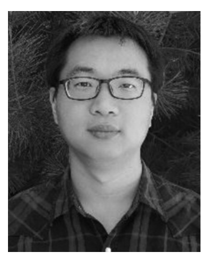


Ying Yang (Senior Member, IEEE) received the Ph.D. degree in control theory from Peking University, Beijing, China, in 2002.

From January 2003 to November 2004, she was a Postdoctoral Researcher with Peking University. From 2005 to 2014, she was an Associate Professor with the Department of Mechanics and Engineering Science, College of Engineering, Peking University, where she has been a Full Professor since 2014. Her research interests include robust and optimal control, nonlinear systems control, numerical analysis, and fault-detection and fault-tolerant systems.

He Li received the B.E. degree in theoretical and applied mechanics from Shanghai University, Shanghai, China, in 2016. She is currently pursuing the Ph.D. degree in dynamical systems and control with Peking University, Beijing, China.

Her current research interests are in the area of gap metrics, manifold optimization, and their role in the design of fault diagnosis systems. She has a special interest in bridging the gap between data-driven methods and model-based methods.



Xianchao Xiu (Member, IEEE) received the Ph.D. degree in operations research from Beijing Jiaotong University, Beijing, China, in 2019.

He is a Postdoctoral Researcher with the State Key Laboratory for Turbulence and Complex Systems, Peking University, Beijing. His current research interests include large-scale sparse optimization, signal processing, machine learning and deep learning, and data-driven fault detection.

Wanquan Liu (Senior Member, IEEE) received the B.S. degree in applied mathematics from Qufu Normal University, Jining, China, in 1985, the M.S. degree in control theory and operation research from the Chinese Academy of Science, Beijing, China, in 1988, and the Ph.D. degree in electrical engineering from Shanghai Jiaotong University, Shanghai, China, in 1993.

He is currently an Associate Professor with the Department of Computing, Curtin University, Perth, WA, Australia. He once held the ARC Fellowship, U2000 Fellowship, and JSPS Fellowship and attracted research funds from different resources over 2.4 million dollars. His current research interests include large-scale pattern recognition, signal processing, machine learning, and control systems.

Dr. Liu is in editorial board for nine international journals.