

Remaining useful life prediction using a nonlinear state-space model under random stress

Mingjun Liu

Xinzhou Normal University

Zengshou Dong

dongzs@tyust.edu.cn

Taiyuan University of Science and Technology

Zhensen Gao

Guangdong University of Technology

Article

Keywords: Remaining useful life (RUL) prediction, Wiener process, Hybrid model, Auxiliary particle filter, Stress model

Posted Date: September 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10809802/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Remaining useful life prediction using a nonlinear state-space model under random stress

Mingjun Liu^{1,2,4}, Zengshou Dong^{3,4*}, Zhensen Gao⁵

¹Department of Electronics, Xinzhou Normal University, Xinzhou, 034000, China

²Intelligent Perception and Fault Prediction Engineering Center of Xinzhou Normal University, Xinzhou, 034000, China

³School of Electronic and Information Engineering, Taiyuan University of Science and Technology, Taiyuan, 030024, China

⁴Shanxi Province Industrial Digitization and Digital Asset Technology Innovation Centre, Taiyuan, 030024, China

⁵School of Information Engineering, Guangdong University of Technology, Guangzhou, 510006, China

*Correspondence and requests for materials should be addressed to Zengshou Dong (email: dongzs@tyust.edu.cn).

Abstract

The harsh operating environment of large slewing bearings subjects the bearing raceways to variable loading conditions, leading to progressive performance degradation and reduced accuracy in remaining useful life prediction. In this study, a nonlinear state-space prediction model is developed to characterise the random effects of multiple degradation features on different degradation states under random raceway stress. An adaptive enhanced auxiliary particle filter algorithm is introduced to improve state estimation accuracy by mitigating the sample degeneracy and scarcity associated with conventional auxiliary particle filter algorithms. Bayesian hyperparameter optimisation is employed to identify the random-stress parameters, and the proposed framework is evaluated based on vibration, temperature, and torque measurements collected from a wind-turbine slewing bearing. Comparative results demonstrate that the joint modelling of random stress and feature–state interactions improves Remaining useful life (RUL) prediction accuracy and provides more reliable support for condition-based maintenance of wind-turbine slewing bearings.

Keywords: Remaining useful life (RUL) prediction; Wiener process; Hybrid model; Auxiliary particle filter; Stress model

1 Introduction

Slewing bearings play a critical role in industrial applications, particularly in wind power generation. They support rotating components and withstand dynamic loads, making their reliable operation essential for maintaining system performance and operational efficiency. However, during operation, wind-turbine slewing bearings are frequently exposed to harsh environmental conditions, emergencies, and random load variations. These factors can promote uneven bearing wear and system degradation, ultimately leading to equipment failure, economic losses, and safety hazards^[1]. To address these challenges, prognostics and health management technologies^[2,3] and RUL prediction^[4,5] have been adopted as key tools for assessing the condition of wind-turbine slewing bearings and supporting timely maintenance decisions.

Existing RUL prediction strategies can be broadly classified into physics-based, data-driven, and hybrid approaches^[6]. For the physics-based approach, Lundbery^[7] established the classical rolling-contact fatigue framework. Building on Lundbery's research, Ioannides^[8] developed a model correlating rolling-contact fatigue life with the reliability of bearings. Zaretsky^[9] subsequently refined the Lundbery model by adopting the maximum shear stress as the characteristic stress, while neglecting the influence of stress depth on bearing service life. Zhou^[10,11] established a load-bearing life prediction model based on fracture mechanics, which accounts for crack initiation and propagation. The reliability of physics-based RUL prediction depends largely on the accuracy of the underlying degradation model. However, for large slewing bearings, geometric complexity, variable loading conditions, and limited access to material and contact parameters make online model calibration difficult and expensive. Moreover, establishing dedicated experimental test platforms is often prohibitively expensive. These challenges restrict experimental validation and hinder the practical application of physics-based RUL prediction.

Unlike physics-based approaches, data-driven methods derive life prediction models directly from monitored degradation data, without relying entirely on physical failure mechanisms. These methods can be broadly categorised into artificial intelligence-based approaches and stochastic process-based statistical methods^[12]. Li^[13] proposed an RUL prediction method based on a deep convolutional neural network. Liu^[14] proposed a joint-loss convolutional neural network for simultaneous fault diagnosis and RUL prediction. Yang et al.^[15] proposed a dual convolutional neural network, the effectiveness of which was demonstrated through four bearing degradation

experiments, achieving superior RUL prediction performance. Zhang^[16] proposed an RUL prediction model based on a dynamic-window neural network capable of adaptively determining the window size, detecting degradation transition points, and performing dynamic prediction. However, these methods require substantial quantities of annotated training data, which are often difficult to obtain in practical industrial settings, thereby posing significant challenges for real-world applications.

Stochastic-process models provide a probabilistic RUL distribution, enabling explicit uncertainty quantification^[17,18]. Among these models, the Gamma and inverse-Gaussian processes impose monotonic degradation paths, whereas the Wiener process captures both upward trends and local reversals. This flexibility makes the Wiener process particularly well suited for modelling the nonlinear and non-monotonic degradation behaviour of wind-turbine slewing bearings^[19]. Si^[20] developed a recursive-filtering Wiener model in which the drift parameter was updated online. In addition, a related nonlinear diffusion formulation with a closed-form solution for RUL prediction was reported by Si et al^[21].

However, most of the previous studies have not explicitly considered the influence of random stress on the degradation process. Li et al.^[22] verified the feasibility of RUL prediction using an Elman neural network through experiments on lithium batteries under vibration stress. Duan et al.^[23] investigated the interaction between sliding motion resistance induced by prestress and friction and the multi-axis inertial forces generated by impact and vibration loads, subsequently establishing a physical model of interlayer sliding. Li et al.^[24] investigated the effects of stress and measurement errors on life prediction based on the accelerated degradation test. Shang et al.^[25] proposed a fatigue damage model that accounts for average stress conversion and cumulative environmental damage. Although considerable progress has been made in RUL prediction based on a wide range of methods and modelling approaches, these studies mainly examined the effects of specific stress factors on equipment performance, with limited consideration of the influence of random stress on the degradation process. Random stress is attributed to the combined effects of multiple factors, making its influence on equipment wear and fatigue more complex than that of individual stress factors.

Given that random stress contributes to the degradation of wind-turbine deflection bearings, a novel degradation model incorporating a random-stress model is developed to improve the accuracy and reliability of RUL prediction for wind-turbine rotary

bearings. In addition, to provide a more comprehensive representation of the operating condition, the proposed model integrates features extracted from multiple sensor signals. The main methodological contributions of this study are summarised as follows:

(1) To account for the random stresses experienced by large slewing bearings during operation, a hybrid degradation model is developed by incorporating a random-stress model into the Wiener process.

(2) A nonlinear state-space prediction model is established to characterise the random effects of multiple degradation features on different degradation states.

(3) An adaptive enhanced auxiliary particle filter (APF) algorithm is proposed to improve state estimation accuracy by mitigating the sample degeneracy and scarcity associated with conventional APF algorithms.

The remainder of the paper is organised as follows: Section 2 formulates the random-stress state-space model; Section 3 presents the enhanced APF algorithm; Section 4 describes the parameter estimation procedure and derives the RUL expressions; Section 5 reports the case study and comparative results; and Section 6 concludes the study.

2 Establishment of a State-Space Model Under Random Stress

Owing to their large size and complex operating conditions, wind-turbine slewing bearings are frequently subjected to significant alternating random stresses that accelerate system degradation. In addition, features extracted from different sensors exhibit varying sensitivities to different degradation states. To capture these characteristics, a nonlinear state-space model under random stress is developed for RUL prediction of wind-turbine slewing bearings (Fig. 1).

The general form of the state-space model is given by the following expression:

$$\begin{cases} x_k = x_{k-1} + \Delta\Lambda(t_{k-1}) + \omega_{k-1} \\ y_k = E \cdot x_k + v_k \end{cases} \quad (1)$$

where x_k and y_k denote the degradation state at time t_k and the characteristics of the monitoring data, respectively. ω_{k-1} and v_k represent the state transition noise and measurement noise, respectively. Given that the degradation process of the wind-turbine slewing bearing exhibits nonlinear and non-monotonic characteristics, a power function is adopted as the drift function, and $\Delta\Lambda(\alpha, t_{k-1}) = \alpha(t_k^r - t_{k-1}^r)$, where α denotes the drift coefficient. In addition, $\omega_{k-1} \sim N(0, \sigma_B^2 B(\Delta t_k))$, $v_k \sim N(0, \delta_e^2)$, where

σ_B denotes the diffusion coefficient, $B(t)$ represents the standard Brownian motion, and E represents the measurement matrix.

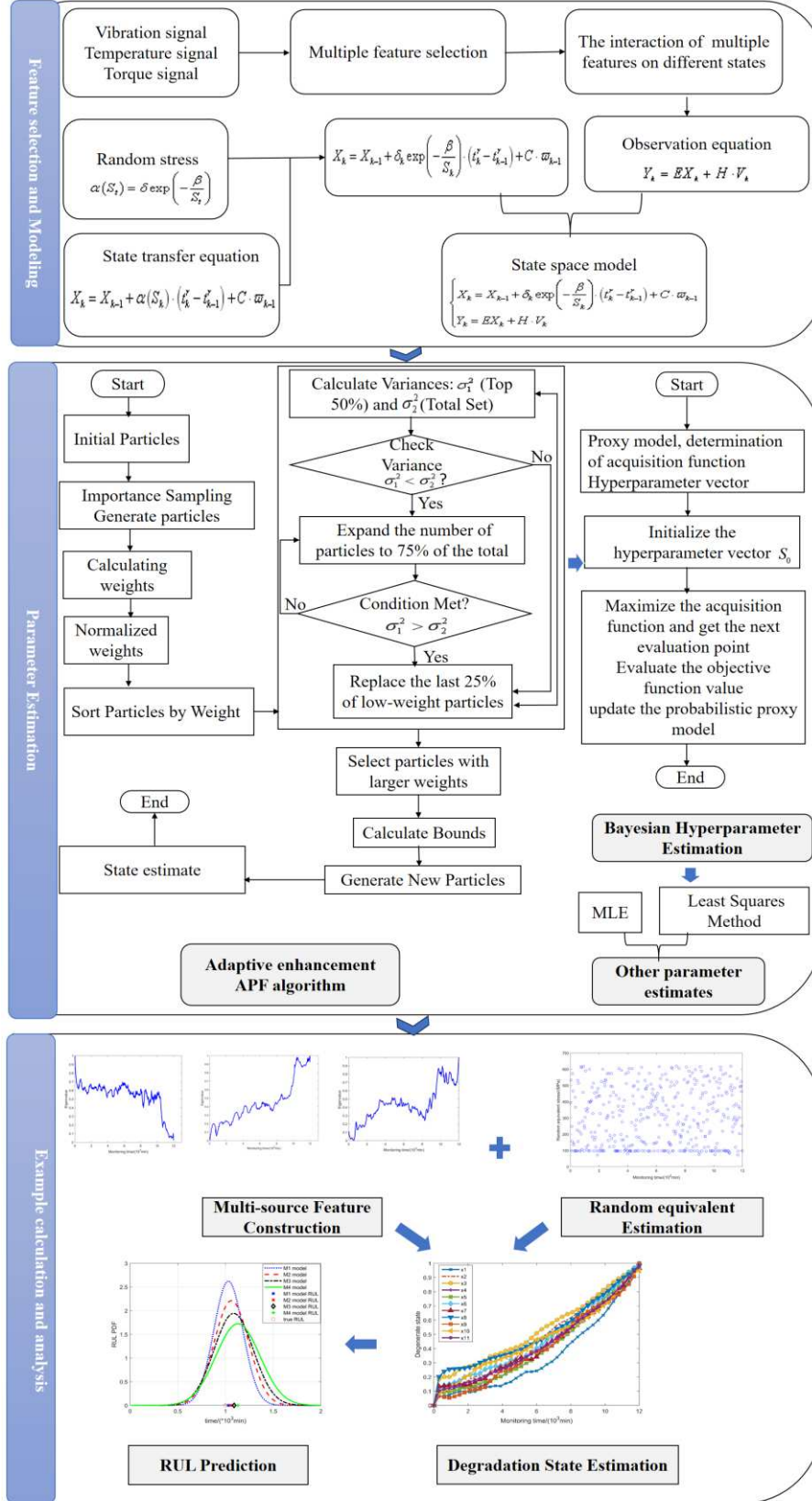


Fig. 1. Flowchart of the proposed RUL prediction method

The Arrhenius acceleration model provides a theoretical basis for describing the relationship between the degradation rate and random stress. Inspired by previous work^[26], the Arrhenius acceleration model is incorporated into the degradation model. Furthermore, to account for the influence of multiple degradation characteristics on different degradation states, a nonlinear state-space model is developed for wind-turbine slewing bearings to improve RUL prediction. This modelling framework provides a more realistic representation of the degradation process. The resulting state transfer equation is expressed as follows:

$$X_k = X_{k-1} + \alpha(S_k) \cdot (t_k^r - t_{k-1}^r) + C \cdot \varpi_{k-1} \quad (2)$$

The degradation state corresponding to feature i at t_k is denoted as $x_{i,k}$, i.e. $X_k = (x_{1,k}, x_{2,k}, \dots, x_{n,k})$. n denotes the total number of features extracted from the data collected by a single sensor. $\alpha(S_k)$ represents the effect of random stress on degradation, corresponding to the degradation rate at time t_k . Furthermore, S_k represents accelerated stress, $B(t) = [B_1(t), B_2(t), \dots, B_n(t)]$. Different values of $B_i(t)$ correspond to different diffusion processes, where $\sigma_B^2 = [\sigma_{B,1}^2, \sigma_{B,2}^2, \dots, \sigma_{B,n}^2]'$, $\varpi_{k-1} = (\omega_{1,k-1}, \omega_{2,k-1}, \dots, \omega_{n,k-1})$. For computational convenience, let C represent the hidden correlation between multiple degenerate states. $c_{i,j} \sigma_{B_{i,j}}^2, (i \neq j)$ denotes the covariance between degenerate states i and j , where $c_{i,j} = 1, (i = j)$, $c_i = (c_{i,1}, c_{i,2}, \dots, c_{i,n})$, $c_i \varpi_{k-1} \sim N(0, c_i \sigma_B^2 B(\Delta t_k))$, $C \cdot \sigma_B^2 B(\Delta t) \sim N(0, \Sigma)$, and Σ represents the diffusion coefficient matrix.

The degradation rate $\alpha(S_t)$ under random stress varies with the applied random stress. Therefore, the drift coefficient is modelled as a function of random stress, as follows:

$$\alpha(S_t) = \delta f(S_t; \Phi) \quad (3)$$

where δ denotes the random coefficient, $\delta = (\delta_1, \delta_2, \dots, \delta_n)'$; δ_n represents the random coefficient of the n -th feature, which follows the Gaussian distribution, $\delta \sim N(\mu_\delta, \sigma_\delta^2)$; n denotes the total number of features; $f(S_t; \Phi)$ represents the degradation coefficient; and Φ represents the degradation parameter. Given the individual differences of the system, the drift coefficient follows a normal distribution, $\alpha(S_t) \sim N[\mu_\delta f(S_t; \Phi), \sigma_\delta^2 f^2(S_t; \Phi)]$. Considering that the applied stress varies randomly, $S_t \sim N(\mu_{S_k}, \sigma_{S_k}^2)$ is assumed.

The Arrhenius model is a widely applied acceleration model, expressed as follows:

$$\alpha(S_t) = \delta \exp\left(-\frac{\beta}{S_t}\right) \quad (4)$$

Substituting Eq. (4) into Eq. (2) yields the expression for the state transfer function under random stresses as follows:

$$Y_k = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n,1} & c_{n,2} & \cdots & c_{n,n} \end{bmatrix} X_k = \begin{bmatrix} c_{2,1} & c_{2,2} & \cdots & c_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n,1} & c_{n,2} & \cdots & c_{n,n} \end{bmatrix} X_{k-1} + \delta_k \exp\left(-\frac{\beta}{S_k}\right) \cdot (t_k^r - t_{k-1}^r) + C \cdot \varpi_{k-1} \quad (5)$$

where $\delta_k = (\delta_{1,k}, \delta_{2,k}, \dots, \delta_{n,k})$, $\Sigma = \begin{bmatrix} c_{1,1}^2 \sigma_{B,1}^2 & c_{1,2}^2 \sigma_{B,2}^2 & \cdots & c_{1,n}^2 \sigma_{B,n}^2 \\ c_{2,1}^2 \sigma_{B,1}^2 & c_{2,2}^2 \sigma_{B,2}^2 & \cdots & c_{2,n}^2 \sigma_{B,n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ c_{n,1}^2 \sigma_{B,1}^2 & c_{n,2}^2 \sigma_{B,2}^2 & \cdots & c_{n,n}^2 \sigma_{B,n}^2 \end{bmatrix}$.

According to the properties of the Wiener process, the following equations are obtained:

$$\Delta X_k \sim N(\mu, \Sigma_1), \quad \mu = \mu_\delta \cdot \exp\left(-\frac{\beta}{S_k}\right) \Psi_k, \quad \Sigma_1 = \sigma_\delta^2 \left(\exp\left(-\frac{\beta}{S_k}\right) \right)^2 \Psi_k \Psi_k' + \Sigma T. \text{ Let } \Psi_k = (\Delta \Lambda(t_{k,1}), \Delta \Lambda(t_{k,2}), \dots, \Delta \Lambda(t_{k,n})).$$

The observation equation is established as follows:

$$Y_k = EX_k + H \cdot V_k \quad (6)$$

where $Y_k = (y_{1,k}, \dots, y_{n,k})$, E represents the observation transfer matrix, and observation errors $V_k = \{v_{1,k}, v_{2,k}, \dots, v_{n,k}\}$, $V_k \sim N(0, \Sigma_2)$ are independent of each other. The parameter matrix H denotes the correlation matrix between different observation data, $H \cdot V_k = \{H_1 \cdot v_{1,k}, H_2 \cdot v_{2,k}, \dots, H_n \cdot v_{n,k}\}$, $H \cdot V_k \sim N(0, \Sigma_3)$. Meanwhile, given the influence of multiple features on different degradation states, the observation transfer matrix E is assumed to be a real symmetric matrix.

Assuming n sets of features, their corresponding observation functions are integrated into multi-source observation functions:

$$\begin{bmatrix} y_{1,k} \\ y_{2,k} \\ \vdots \\ y_{n,k} \end{bmatrix} = \begin{bmatrix} e_{1,1} & e_{1,2} & \cdots & e_{1,n} \\ e_{2,1} & e_{2,2} & \cdots & e_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ e_{n,1} & e_{n,2} & \cdots & e_{n,n} \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \\ \vdots \\ x_{n,k} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} + H \cdot \begin{bmatrix} v_{1,k} \\ v_{2,k} \\ \vdots \\ v_{n,k} \end{bmatrix} \quad (7)$$

where $y_k = (y_{1,k}, \dots, y_{n,k})$ denotes the characteristic vector of the n -th group of observation data.

Given that

Therefore,

where $h_{i,j} \sigma_i^2, (i \neq j)$ denotes the covariance of the measurement noise between features i and j , let $h_{i,j} = 1, (i = j)$. According to the outlined formulation, the state-space model can be expressed as follows:

$$\begin{cases} X_k = X_{k-1} + \delta_k \exp\left(-\frac{\beta}{S_k}\right) \cdot (t_k^r - t_{k-1}^r) + C \cdot \varpi_{k-1} \\ Y_k = EX_k + H \cdot V_k \end{cases} \quad (8)$$

3 Analysis of the Adaptive Enhanced APF Algorithm

The APF algorithm is designed to mitigate the effects of sample degeneracy during particle estimation, thereby improving state estimation accuracy. The proposed adaptive enhanced APF algorithm compares particle variances to adaptively identify particles with higher weights for replacement of those with lower weights, thereby maintaining particle diversity while increasing particle weights.

Assume a particle set containing m particles, denoted by $\{x_{i,k}, w_{i,k}, i = 1, 2, \dots, m\}$. For the posterior probability density function $p(x_k | y_{1:k})$, direct sampling remains computationally intractable. Therefore, an alternative importance sampling distribution is introduced: $q(x_k | y_{1:k})$.

The particle weights are computed according to the particle weight update equation proposed by Li^[27], and the particles are ranked in descending order according to their weights, $W = \{w_{i,k}\}$, where $w_{1,k} \geq w_{2,k} \geq \dots \geq w_{i,k}$.

During adaptive particle number selection, either the variance or the likelihood function of the particle set is typically used as the threshold selection criterion. In this study, the particle variance is adopted as the optimisation criterion. A smaller variance indicates a more concentrated particle distribution and, consequently, higher state estimation accuracy. Therefore, the objective function is defined as the variance of the particle set, as follows:

$$f(x) = \frac{1}{m} \sum_{i=1}^m (x_i - \bar{x})^2 \quad (9)$$

where x denotes the particle set, and $\bar{x}$ denotes the mean of the particle set.

Following the principle of a one-dimensional search method, the variance σ_1^2 of the first 50% of the particles is calculated. Let σ_2^2 denote the variance of the entire particle set. Through the comparison of these variances, the range of replacing particles is determined, after which the particles with higher weights are redistributed. With the particle corresponding to the maximum weight $w_{1,k}$ given by $x_{i,k}$, the upper and lower limits are calculated as follows:

$$\begin{cases} L_b = x_{i,k} - \sigma \\ U_b = x_{i,k} + \sigma \end{cases} \quad (10)$$

Subsequently, the particles with lower weights are replaced by newly generated particles with higher weights. To preserve particle diversity, the new particles are generated according to the following formula:

$$x_{j,k} = L_b + (U_b - L_b)\eta \quad (11)$$

where $\eta \in [0,1]$ represents a random number, reflecting the diversity of particles.

The particles are subsequently re-ranked according to their weights to obtain $W' = \{w'_{i,k}\}$, where $w'_{1,k} \geq w'_{2,k} \geq \dots \geq w'_{i,k}$.

Table 1 summarises the proposed adaptive enhanced APF algorithm.

Table 1. Adaptive enhanced APF algorithm

Step 1. Initialisation: Initialise the particle set $x_{i,0} \sim p(x_0), i = 1, 2, \dots, m$

Step 2. Importance sampling: $k = 1 : m$ Generate a new set of importance particles, $x_{i,k} \sim q(x_{i,k} | x_{i,k-1}, y_{0:k})$

Step 3. Weight calculation: $w_{i,k} = w_{i,k-1} \frac{p(y_k | x_{i,k}) p(x_{i,k} | x_{i,k-1})}{q(x_{i,k} | x_{i,k-1}, y_{1:k})}$

Step 4. Weight normalisation: $\bar{w}_{i,k} = w_{i,k} / \sum_{i=1}^m w_{i,k}$

Step 5. Adaptive particle enhancement:

Sort the particle weights to obtain $W = \{w_{i,k}\}$, where,

$$w_{1,k} \geq w_{2,k} \geq \dots \geq w_{i,k}$$

(1) Calculate the variance σ_1^2 of the first 50% of the particles. Let σ_2^2 denote the variance of the entire particle set.

If $\sigma_1^2 < \sigma_2^2$

Then increase the number of particles to 75% of the total

If $\sigma_1^2 > \sigma_2^2$

Then replace the last 25% of low-weight particles

Else

Continue increasing the number of particles until $\sigma_1^2 > \sigma_2^2$

End

(2) Select particles with larger weights

Let the particle corresponding to $w_{1,k}$ be denoted by $x_{i,k}$, calculate the upper and lower limits $L_b = x_{i,k} - \sigma$; $U_b = x_{i,k} + \sigma$

Generate new particles according to the following formula to replace the low-weight particles $x_{j,k} = L_b + (U_b - L_b) \eta_m$

Step 6. Output: State estimate $\hat{x}_k = \sum_{i=1}^m w_{i,k} x_{i,k}$

The particle filter effect can be evaluated via several methods^[28]. In this study, the effective sample size (ESS) is adopted to assess the effectiveness of the proposed method, defined as follows:

$$ESS(k) = \frac{\left(\sum_{i=1}^m w_{i,k} \right)^2}{\sum_{i=1}^m (w_{i,k})^2} \times 100\% \quad (12)$$

where $w_{i,k}$ denotes the weight of the importance particle at the k -th time step. When most particles are located within a given high-probability region (i.e., $w_{i,k} \geq 1/m$), the ESS reaches its maximum value (i.e., 100%). Conversely, a decrease in the ESS indicates a higher degree of sample degeneracy.

4 Parameter Estimation and RUL Prediction

4.1 Bayesian Hyperparameter Estimation

Maximum Likelihood Estimate (MLE) and the least squares method are employed for adaptive estimation of unknown parameters, $\theta = [\mu_\eta, \sigma_\eta^2, \sigma_B^2, r, \sigma_e^2, C, E, H, S_k, \beta]$. At time t_k , the MLE of the parameter is expressed as follows:

$$\hat{\theta}^{(k+1)} = \max \left\{ \ln f(y_{1:k}, \hat{x}_k; \theta) \right\} \quad (13)$$

where $Y_{1:k} = (Y_{1,1:k}, Y_{2,1:k}, \dots, Y_{n,1:k})$ represents the characteristic of the observed data, $\hat{X}_k = (\hat{x}_{1,k}, \hat{x}_{2,k}, \dots, \hat{x}_{m,k})$ denotes the estimated state, and $f(Y_{1:k}, \hat{X}_k; \theta)$ represents the joint probability density function of multiple features extracted from the observed data and the system degradation state.

Given $S_i \sim N(\mu_{s_k}, \sigma_{s_k}^2)$, direct application of MLE to the acceleration model is computationally challenging. In this study, the Bayesian hyperparameter method is employed to estimate $A(\mu_{s_k}, \sigma_{s_k}^2)$. To reduce the computational complexity of the model, a Gaussian process is adopted as the surrogate model. Let the test accuracy data Z serve as a random observation, such that $z = f(S_k) + \varepsilon$, where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ denotes independent and identically distributed noise.

For a model output value of 0, the joint distribution expression of known data $H_i = (S_i, z_i)$, ($i = 1, 2, \dots, t$), and a new observation point $(\hat{X}_{i+1}, p_{i+1})$ is as follows:

$$[p_{1:t+1}] \sim N \left(0, \begin{bmatrix} B + \sigma_\varepsilon^2 \cdot I & B_1' \\ (B_1')^T & B_1(S_{i+1}, S_{i+1}) \end{bmatrix} \right) \quad (14)$$

where B_1 denotes the covariance function of $S * S$, $B_1' = [B_1(S_1, S_{i+1}), \dots, B_1(S_i, S_{i+1})]^T$, I denotes the identity matrix, and the Gram matrix is defined as follows:

$$B = \begin{bmatrix} B_1(S_1, S_1) & \cdots & B_1(S_1, S_i) \\ \vdots & \ddots & \vdots \\ B_1(S_i, S_1) & \cdots & B_1(S_i, S_i) \end{bmatrix} \quad (15)$$

The posterior distribution of y_{i+1} is as follows:

$$p(Y_{i+1} | Y_{1:t}, S_{1:t+1}) = N(\mu_t(S_{t+1}), \sigma_t^2(S_{t+1})) \quad (16)$$

Thus, the posterior mean and variance are obtained as follows:

$$\mu_t(S_{t+1}) = (B_1')^T (B + \sigma_\varepsilon^2 I)^{-1} y_{1:t} \quad (17)$$

$$\sigma_t^2(S_{t+1}) = B_1(S_{t+1}, S_{t+1}) - (B_1')^T (B + \sigma_\varepsilon^2 I)^{-1} B_1' \quad (18)$$

Generally, the automatic correlation is employed to determine the squared exponential function as a Gaussian regression:

$$B(S, S') = \theta_0 \exp\left\{-\frac{1}{2} \kappa^2(S, S')\right\} \quad (19)$$

$$(20)$$

The next sample point is selected by maximising the acquisition function Δ , which is expressed as follows:

$$\kappa^2(S, S') = \sum_{num=1}^N (S, S')^2 / \theta_{num}^2 \quad (21)$$

The EI (expected improvement) enhancement function is adopted as the acquisition function, as follows:

$$\Delta(x) = \begin{cases} \left(\mu(S) - f(S^+) + \sigma(S) \right) \Phi\left(\frac{\mu(S) - f(S^+)}{\sigma(S)} \right), & \sigma(S) > 0 \\ 0, & \sigma(S) = 0 \end{cases} \quad (22)$$

where $f(S^+)$ denotes the current maximum value.

Table 2 summarises the hyperparameter estimation process.

Table 2. Hyperparameter estimation process

Step 1. Input: Proxy model and acquisition function
Output: Hyperparameter vector
Step 2. Initialise the hyperparameter vector S_0

Step 3. For $i = 1, 2, \dots, t$

Maximise the acquisition function to determine the next evaluation point: $S_{t+1} = \arg \max \Delta_t(S; N)$

Evaluate the objective function value $y_{t+1} = f(S_{t+1}) + \varepsilon_{t+1}$

Integrate data, $N_i = N \cup (S_{i+1}, y_{i+1})$, and update the probabilistic proxy model.

End

4.2 Estimation of Other Parameters

The joint log-likelihood function of the system degradation state and the observed data is expressed as follows:

$$\ln f(y_{1:k} | \theta) = \ln f(\hat{x}_k | \theta) + \ln f(y_{i,1:k} | \hat{x}_k, \theta) \quad (23)$$

Assuming that the multi-sensor observations are mutually independent yields the following:

$$f(y_{1:k} | \hat{x}_k; \theta) = \prod_{i=1}^n f(y_{i,1:k} | \hat{x}_k, \theta) \quad (24)$$

Substituting Eq. (24) into Eq. (23) yields the following:

$$\begin{aligned} \ln f(\hat{x}_k, y_{1:k}; \theta) &= \ln f(\hat{x}_k; \theta) + \ln \prod_{i=1}^n f(y_{i,1:k} | \hat{x}_k; \theta) \\ &= \ln f(\hat{x}_k | \theta) + \sum_{i=1}^n \ln f(y_{i,1:k} | \hat{x}_k; \theta) \end{aligned} \quad (25)$$

The Bayesian hyperparameter estimation of $\mu_{s_k}, \sigma_{s_k}^2$, and random stress S_k is used to conveniently estimate other parameters:

$$g(t_k; \theta') = \exp\left(-\frac{\beta}{S_k}\right) \cdot (t_k^r - t_{k-1}^r) \quad (26)$$

where parameter $\theta' = [\beta, r]$. Subsequently, $\hat{\theta} = [\hat{\mu}_\eta, \hat{\sigma}_\eta^2, \hat{\sigma}_B^2, \hat{\sigma}_e^2, \hat{C}, \hat{D}, \hat{E}, \hat{H}, \hat{\beta}]$ is estimated using Eq. (8), as follows:

$$\begin{aligned} \ln f(\hat{x}_k | \theta) &= \sum_{z=1}^k \left\{ -\frac{1}{2} \ln |\Sigma_{1,i}| - \frac{1}{2} (x_{i,z} - x_{i,z-1} - \mu_\eta g(t_k; \theta'))' \Sigma_{1,i}^{-1} (x_{i,z} - x_{i,z-1} - \mu_\eta g(t_k; \theta')) \right\} \\ &= -\frac{k}{2} \ln |\Sigma_{1,i}| - \frac{\sum_{z=1}^k (x_{i,z} - x_{i,z-1} - \mu_\eta g(t_k; \theta'))' \Sigma_{1,i}^{-1} (x_{i,z} - x_{i,z-1} - \mu_\eta g(t_k; \theta'))}{2} \end{aligned} \quad (27)$$

$$\begin{aligned}
\ln f(y_{1:k} | x_k; \theta) &= \ln \left\{ \prod_{z=1}^k \frac{1}{\sqrt{2\pi(\Sigma_{3,i})^2}} \exp \left[-\frac{(y_{i,z} - E_i x_{i,z})^2}{2(\Sigma_{3,i})^2} \right] \right\} \\
&= -\frac{k}{2} \ln 2\pi(\Sigma_{3,i})^2 - \frac{1}{2(\Sigma_{3,i})^2} \sum_{z=1}^k (y_{i,z} - E_i x_{i,z})^2
\end{aligned} \tag{28}$$

where x_k and x_{k-1} represent the state values at time $k-1$ and time k , respectively. Combining Eqs. (27) and (28) yields the following:

$$\begin{aligned}
\ln f(x_k, y_{1:k}; \theta) &= \ln f(x_k; \theta) + \ln \prod_{i=1}^n f(y_{i,1:k} | x_k; \theta) \\
&= -\frac{k}{2} \ln |\Sigma_{1,i}| - \frac{\sum_{z=1}^k (x_{i,z} - x_{i,z-1} - \mu_{\eta} g(t_k; \theta'))' \Sigma_{1,i}^{-1} (x_{i,z} - x_{i,z-1} - \mu_{\eta} g(t_k; \theta'))}{2} \\
&\quad + \sum_{i=1}^n \left(-\frac{k}{2} \ln 2\pi(\Sigma_{3,i})^2 - \frac{1}{2(\Sigma_{3,i})^2} \sum_{z=1}^k (y_{i,z} - E_i \hat{x}_{i,z})^2 \right)
\end{aligned} \tag{29}$$

Given r , the likelihood function $\ln f(x_k, y_{1:k}^{1:n}; \theta)$ is used to estimate the parameter $\hat{\theta}_\tau$ in the state transfer equation:

$$\begin{aligned}
\hat{\mu}_{\eta_k, \tau+1} &= \frac{1}{\sum_{z=1}^k 2 \left(g(t_{\tau+1,i,z}; \theta')' \right) \Sigma_{1,i, \tau+1}^{-1} g(t_{\tau+1,i,z}; \theta')} \sum_{z=1}^k \left[\left(g(t_{\tau+1,i,z}; \theta')' \right) \Sigma_{1,i, \tau+1}^{-1} (\hat{X}_{\tau,i,z} - \hat{X}_{\tau,i,z-1}) \right. \\
&\quad \left. + g(t_{\tau+1,i,z}; \theta') \Sigma_{1,i, \tau+1}^{-1} (\hat{X}_{\tau,i,z} - \hat{X}_{\tau,i,z-1})' \right]
\end{aligned} \tag{30}$$

$$\begin{aligned}
\hat{\Sigma}_{1,i, \tau+1} &= \frac{1}{2k} \sum_{z=1}^k \left[(\hat{X}'_{\tau,i,z} \hat{X}_{\tau,i,z}) - (\hat{X}'_{\tau,i,z} \hat{X}_{\tau,i,z-1}) - \hat{X}'_{\tau,i,z} \mu_{\eta_k, \tau+1} g(t_{\tau+1,i,z}; \theta') - (\hat{X}'_{\tau,i,z-1} \hat{X}_{\tau,i,z}) \right. \\
&\quad \left. + (\hat{X}'_{\tau,i,z-1} \hat{X}_{\tau,i,z-1}) + \hat{X}'_{\tau,i,z-1} \mu_{\eta_k, \tau+1} g(t_{\tau+1,i,z}; \theta') - (\mu_{\eta_k, \tau+1} g(t_{\tau+1,i,z}; \theta'))' \hat{X}_{\tau,i,z} \right. \\
&\quad \left. + (\mu_{\eta_k, \tau+1} g(t_{\tau+1,i,z}; \theta'))' \hat{X}_{\tau,i,z-1} + (\mu_{\eta_k, \tau+1} g(t_{\tau+1,i,z}; \theta'))' \mu_{\eta_k, \tau+1} g(t_{\tau+1,i,z}; \theta') \right]
\end{aligned} \tag{31}$$

$$\begin{aligned}
\hat{\sigma}_{\eta_k, \tau+1}^2 &= \frac{1}{g(t_{\tau+1,i,z}; \theta') g(t_{\tau+1,i,z}; \theta')} \left\{ \frac{1}{2k} \sum_{z=1}^k \left[(\hat{X}'_{\tau,i,z} \hat{X}_{\tau,i,z}) - (\hat{X}'_{\tau,i,z} \hat{X}_{\tau,i,z-1}) - \hat{X}'_{\tau,i,z} \mu_{\eta_k, \tau+1} g(t_{\tau+1,i,z}; \theta') - (\hat{X}'_{\tau,i,z-1} \hat{X}_{\tau,i,z}) \right. \right. \\
&\quad \left. \left. + (\hat{X}'_{\tau,i,z-1} \hat{X}_{\tau,i,z-1}) + \hat{X}'_{\tau,i,z-1} \mu_{\eta_k, \tau+1} g(t_{\tau+1,i,z}; \theta') - (\mu_{\eta_k, \tau+1} g(t_{\tau+1,i,z}; \theta'))' \hat{X}_{\tau,i,z} \right. \right. \\
&\quad \left. \left. + (\mu_{\eta_k, \tau+1} g(t_{\tau+1,i,z}; \theta'))' \hat{X}_{\tau,i,z-1} + (\mu_{\eta_k, \tau+1} g(t_{\tau+1,i,z}; \theta'))' \mu_{\eta_k, \tau+1} g(t_{\tau+1,i,z}; \theta') \right] - \Sigma T \right\}
\end{aligned} \tag{32}$$

$$\hat{\beta}_{k,\tau+1} = -S_{k,\tau+1} \cdot \ln \left(\frac{\mu'_{\eta_k,\tau+1}(t_{\tau,i,z} - t_{\tau,i,z-1})' \Sigma_{1,i,\tau+1}^{-1} (\hat{X}_{\tau,i,z} - \hat{X}_{\tau,i,z-1}) + (\hat{X}_{\tau,i,z} - \hat{X}_{\tau,i,z-1})' \Sigma_{1,i,\tau+1}^{-1} \mu_{\eta_k,\tau+1}(t_{\tau,i,z} - t_{\tau,i,z-1})}{2\mu'_{\eta_k,\tau+1}(t_{\tau,i,z} - t_{\tau,i,z-1})' \Sigma_{1,i,\tau+1}^{-1} \mu_{\eta_k,\tau+1}(t_{\tau,i,z} - t_{\tau,i,z-1})} \right) \quad (33)$$

Accordingly,

$$\begin{aligned} \hat{\Sigma}_{i,\tau+1} = & \frac{1}{2kT} \sum_{z=1}^k \left[(\hat{X}'_{\tau,i,z} \hat{X}_{\tau,i,z}) - (\hat{X}'_{\tau,i,z} \hat{X}_{\tau,i,z-1}) - \hat{X}'_{\tau,i,z} \mu_{\eta_k,\tau+1} g(t_{\tau+1,i,z}; \theta') - (\hat{X}'_{\tau,i,z-1} \hat{X}_{\tau,i,z}) \right. \\ & + (\hat{X}'_{\tau,i,z-1} \hat{X}_{\tau,i,z-1}) + \hat{X}'_{\tau,i,z-1} \mu_{\eta_k,\tau+1} g(t_{\tau+1,i,z}; \theta') - (\mu_{\eta_k,\tau+1} g(t_{\tau+1,i,z}; \theta'))' \hat{X}_{\tau,i,z} \\ & \left. + (\mu_{\eta_k,\tau+1} g(t_{\tau+1,i,z}; \theta'))' \hat{X}_{\tau,i,z-1} + (\mu_{\eta_k,\tau+1} g(t_{\tau+1,i,z}; \theta'))' \mu_{\eta_k,\tau+1} g(t_{\tau+1,i,z}; \theta') \right] \\ & - \frac{1}{T} \sigma_{\eta_k,\tau+1} \left(\exp \left(-\frac{\beta}{S_k} \right) \right)^2 \Lambda_{\tau+1,k} \Lambda'_{\tau+1,k} \end{aligned} \quad (34)$$

Given $c_{i,j} = 1, (i = j)$, $C \cdot \sigma_B^2 B(\Delta t) \sim N(0, \Sigma)$, the correlation coefficient matrix

C can be calculated.

Subsequently, with the estimated parameters $(\hat{\theta}_i | r), i = 1, 2, \dots, m$ substituted into the maximum likelihood function (Eq. 29), $\hat{r}$ is estimated as follows:

$$\ln L(\hat{r} | \theta_i) \propto -\frac{k}{2} \ln |\Sigma_{1,i}| - \frac{\sum_{z=1}^k (x_{i,z} - x_{i,z-1} - \mu_{\eta} g(t_k; \theta'))' \Sigma_{1,i}^{-1} (x_{i,z} - x_{i,z-1} - \mu_{\eta} g(t_k; \theta'))}{2} \quad (35)$$

In the measurement equation, the u -th sensor is assumed to extract n distinct features, each exhibiting a different degradation path. External disturbances attributed to the same operating conditions introduce correlated measurement noise among these features. The degradation process associated with each feature corresponds to its own parameters e_k , h_k , and Σ_3 . The correlation among the measurement noises is described by the covariance matrix Σ_1 , and the parameters of the measurement function are estimated in two steps. First, the model parameters e_k , h_k , and Σ_3 are estimated from the extracted features. Second, the covariance matrix Σ_2 is estimated using both the model parameters and the extracted features.

First, the data are smoothed via a local regression method^[106] to reduce measurement noise. The filtered data are presented as follows:

$$\hat{Y}_k = E \hat{X}_k \quad (36)$$

According to the degradation state and the measurement equation, the target optimisation function is formulated as follows:

$$\begin{aligned}\min J_2 &= \min \left[\sum_{i=1}^n (y_{i,k} - \hat{y}_{i,k})^2 \right] \\ &= \min \left[\sum_{i=1}^n (y_{i,k} - e_i \hat{x}_{i,k})^2 \right]\end{aligned}\quad (37)$$

The least squares method is employed to estimate the measurement equation parameter e_k , thereby obtaining the measurement matrix E .

Second, the covariance matrix H containing the correlation information of different measurement points and the variance matrix Σ_2 are computed according to Eq. (8). The dataset $Y_{n,k1,n} = (y_{1,k}, y_{2,k}, \dots, y_{n,k})$ with n features and multiple measurement points at time t_k satisfies the multivariate normal distribution $N(E\Phi_{n,k}, \Sigma_3)$. where $E = \begin{bmatrix} e_{n,n} & e_{n,1} & \dots & e_{n,n} \end{bmatrix}$, $H = \begin{bmatrix} h_{1,n} & h_{2,n} & \dots & h_{n,n} \end{bmatrix}$.

The log-likelihood function of the covariance matrix Σ_3 is given by:

$$L(\Sigma_3|Y) = \sum_{i=1}^n \left(-\frac{k}{2} - \frac{k}{2} \ln |\Sigma_3| - \frac{\sum_{z=1}^k (Y_{i,z} - EX_{i,z}) \Sigma_3^{-1} (Y_{i,z} - EX_{i,z})^T}{2} \right) \quad (38)$$

The maximum likelihood estimate of the covariance matrix Σ_3 is given by:

$$\hat{\Sigma}_{3,\tau+1} = \frac{\sum_{i=1}^n \sum_{z=1}^k (Y_{\tau,i,z} - E_{\tau,i} X_{\tau,i,z}) (Y_{\tau,i,z} - E_{\tau,i} X_{\tau,i,z})^T}{\sum_{i=1}^n k} \quad (39)$$

Given $H\Sigma_2 = \hat{\Sigma}_3$, $h_{i,j} = 1, (i = j)$, Σ_2 , $h_{i,j}, (i \neq j)$, H , and σ_e can be calculated.

4.3 RUL Prediction

For the sensor data of unit m , the fused degradation state is represented by the time series $T^m = (t_1^m, t_2^m, \dots, t_n^m)$, where n denotes the total number of samples. The sequence corresponding to the i -th degradation state is denoted by $\{x_{i,1}^m, x_{i,2}^m, \dots, x_{i,k}^m\}$. Given that the degradation states obtained from different degradation processes are as follows: $X_{1:k} = \{X_1, X_2, \dots, X_k\}$, α and δ_B are treated as independent random variables, such that $\alpha \sim N(\mu_\alpha, \delta_\alpha^2)$.

The total probability formula yields:

$$f_{L_k|X_{a,1:k}}(l_k | X_{a,1:k}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{L_k|\alpha, D_e, X_{a,1:k}}(l_k | \alpha, D_e, X_{a,1:k}) p(D_e | \alpha, X_{a,1:k}) p(\alpha | X_{a,1:k}) dD_e d\alpha \quad (40)$$

$$= E_{\alpha|X_{a,1:k}} \left[E_{D_e|\alpha, X_{a,1:k}} \left[f_{L_k|\alpha, D_e, X_{a,1:k}}(l_k | \alpha, D_e, X_{a,1:k}) \right] \right]$$

Theorem 1^[29]: If $Z \sim N(\mu_z, \delta_z^2)$ and $a, b, c, d, r \in R$, then:

$$E_Z \left[(a' - b'Z') \cdot \exp \left(-\frac{(c' - d'Z')^2}{2\gamma} \right) \right] = \sqrt{\frac{\gamma}{d'^2 \sigma_z^2 + \gamma}} \times \left(a' - b' \frac{d' \sigma_z^2 c' + \mu_z \gamma}{d'^2 \sigma_z^2 + \gamma} \right) \exp \left(-\frac{(c' - d' \mu_z)^2}{2(d'^2 \sigma_z^2 + \gamma)} \right) \quad (41)$$

Combining Theorem 1 with Eq. (40) yields an approximate expression of the PDF

$$f_{L_k|Y_{1:k}}(l_k | Y_{1:k}) \text{ of the RUL: If } D_e \sim N(D - \mu_\Omega, \delta_\Omega^2), \alpha \sim N(\mu_\alpha, \delta_\alpha^2), \text{ then:}$$

$$f(l_k) \cong \frac{1}{\sqrt{2\pi l_k^2 (q'^2 \sigma_\eta^2 + l_k \sigma_B^2)^3}} \left[(D - X_k) (l_k \sigma_B^2 - r g(t_k; \theta') l_k (l_k + t_k)^{r-1} q' \sigma_\eta^2) \right. \quad (42)$$

$$\left. - (q' - r g(t_k; \theta') l_k (l_k + t_k)^{r-1}) \mu_\eta l_k \sigma_B^2 \right] \bullet \exp \left(\frac{-(D - X_k - q' \mu_\eta)^2}{2(q'^2 \sigma_\eta^2 + l_k \sigma_B^2)} \right)$$

where $q' = g(t_k; \theta')$, $a' = c' = D - X_k$, $b' = q' - r l_k (l_k + t_k)^{r-1}$, $d' = q'$.

5 Example Calculation and Analysis

This section presents a numerical analysis and a case study of a wind-turbine slewing bearing to demonstrate the effectiveness of the proposed method, including random stress estimation, state estimation, parameter estimation, and RUL prediction. The experimental dataset comprises vibration, temperature, and torque measurements collected from a wind-turbine slewing bearing and reported by Feng^[30]. During operation, failure of a wind-turbine slewing bearing is mainly attributed to local raceway deformation and rolling-contact fatigue spalling. Accordingly, the influence of contact stress between the raceway and the rolling elements on the degradation process is investigated. According to the literature^[31], the unit type C3D8I is adopted, and the swept meshing method is employed to model the contact between the ball and the raceway. Tables 3 and 4 present the material parameters of the slewing bearing. Given these parameters, a finite element model is developed in ABAQUS to perform the stress analysis. Figure 2 shows the simulated raceway force distribution.

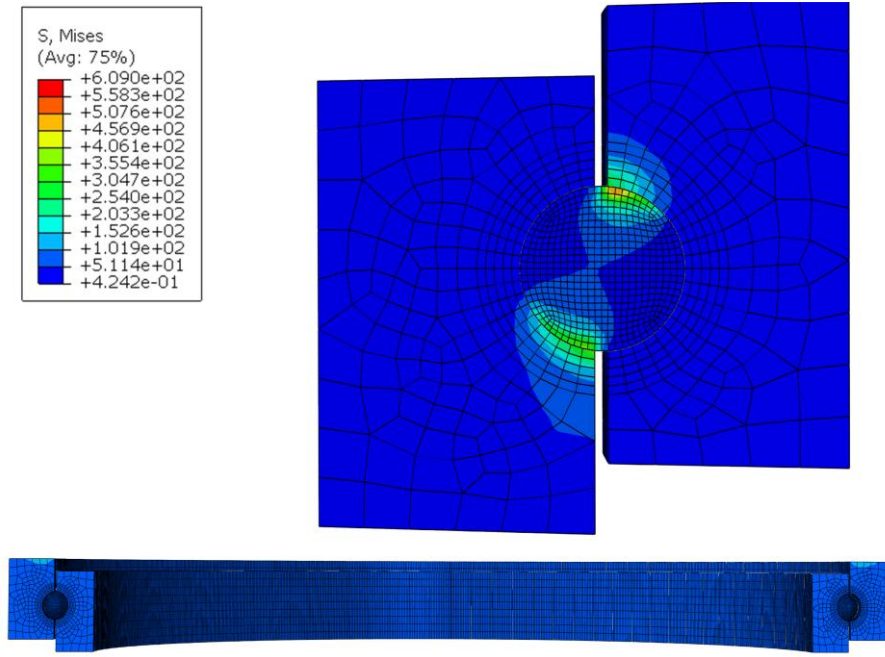


Fig. 2. Slewing bearing stress gradient

Table 3. Geometric properties of the tested slewing bearing

Raceway diameter (mm)	Ball diameter (mm)	Axial force(kN)	Radial force (kN)	Overturning moment (kN·m)	Speed range(r/ min)	Number of balls
730	22	0-750	0-101.8	0-880	0.5-5	91

Table 4. Material properties of the tested slewing bearing

	Materials	Young's modulus (GPa)	Poisson's ratio	Cyclic strength coefficient (MPa)	Cyclic strain hardening index	Fatigue strength coefficient (MPa)	Fatigue elongation coefficient
Ball	GCr15	207	0.3	1173	0.0932	1511	1.3943
Race way	42CrMo	212	0.3	5808	0.1550	2425	0.1167

The minimum and maximum equivalent stresses of the rolling ball and the raceway contact area are 51.2 MPa and 609 MPa, respectively, at which the axial force and overturning moment are coupled with the maximum position (Fig. 2). Before parameter estimation and RUL prediction, the parameters in Eq. (4) are estimated from the equivalent stress range using Bayesian hyperparameter optimisation. The estimated parameters are then used to determine the random stress values at different monitoring times (Fig. 3).

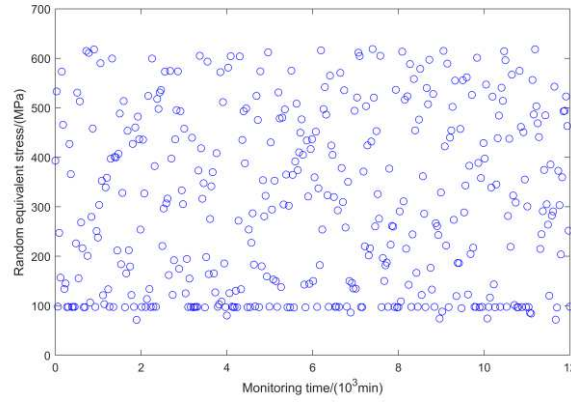
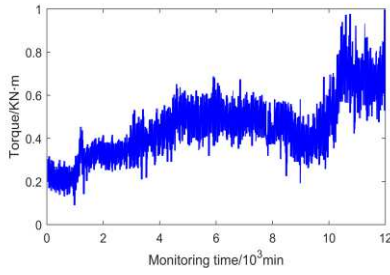


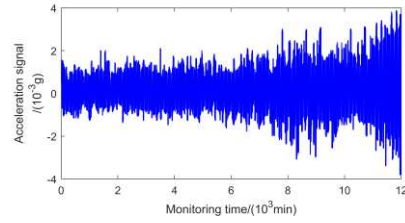
Fig. 3. Random equivalent stress of the slewing bearing

Before parameter $\theta = [\mu_\eta, \sigma_\eta^2, \sigma_B^2, r, \sigma_e^2, C, E, H, S_k, \beta]$ estimation and RUL prediction, Bayesian hyperparameter optimisation is employed to estimate the parameter $(\mu_{s_k}, \sigma_{s_k}^2)$ in Eq. (4) based on the equivalent stress range. The estimated parameter is subsequently used to determine the stress values at different monitoring times. These values characterise the stress conditions of the raceway throughout the monitoring period and provide essential data for the subsequent RUL prediction.

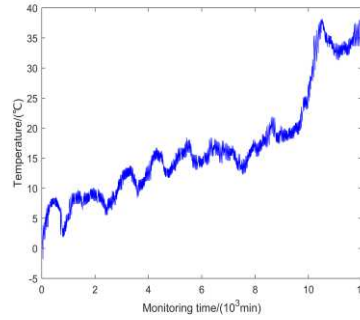
To validate the proposed method, the original vibration, temperature, and torque data reported by Feng^[30] are used. The torque, vibration, and temperature data are integrated for model validation and are denoted by a_{51} , a_{52} , and a_{53} , respectively. Figure 4 presents the fitted torque, vibration, and temperature data.



(a) Driven torque data



(b) Vibration data



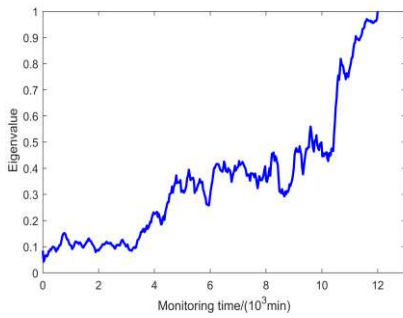
(c) Temperature data

Fig. 4. Fitted torque, vibration, and temperature data

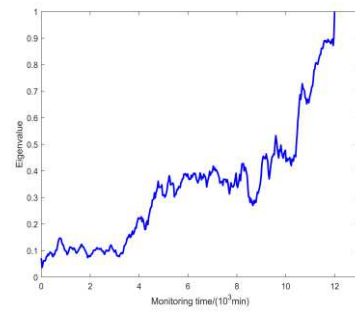
According to the feature extraction method proposed by Liu^[32], the square root amplitude, average value, root mean square value, wavelet packet features, maximum value, pulse index, margin index, and peak index are selected as sensitive features for the vibration data. Feature extraction is subsequently performed on the torque and temperature signals, and the corresponding Spearman correlation coefficients are calculated. Table 5 presents the results.

Table 5. Time-domain feature indices and Spearman correlation coefficients

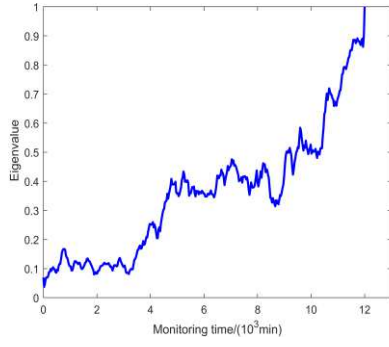
Feature	Feature expression	Torque correlation coefficient	Temperature signal correlation coefficient
Maximum value	$y_{\max} = \max\{ Y_i \}$	0.92	0.91
Peak-to-peak value	$y_{p-p} = \max(Y_i) - \min(Y_i)$	0.84	0.90
Variance	$y_{\sigma^2} = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2$	0.85	0.89
Root mean square	$y_{rms} = \sqrt{\frac{1}{n} \sum_{i=1}^n Y_i^2}$	1.00	0.91
Kurtosis	$y_K = \frac{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^4}{\left(\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2\right)^2}$	0.60	0.89



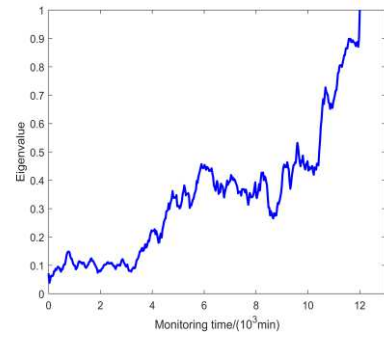
(a) Normalization of x_r



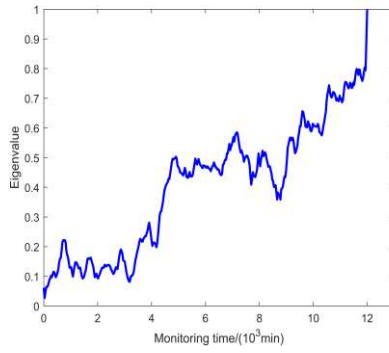
(b) Normalization of x_{mean}



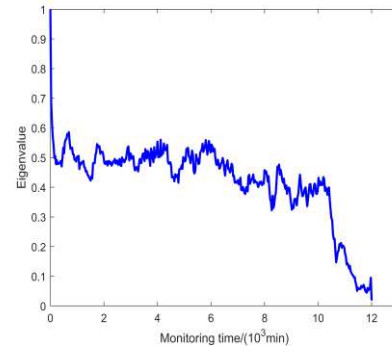
(c) Normalization of x_{rmx}



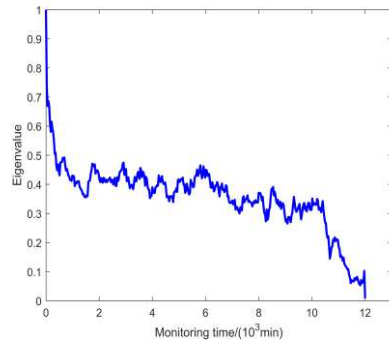
(d) Feature normalization of the wavelet packet



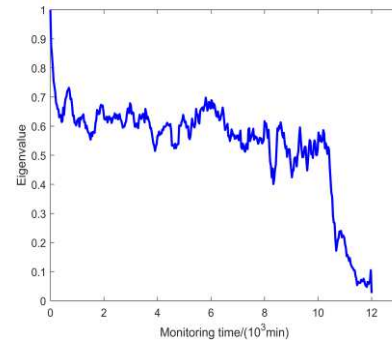
(e) Normalization of x_{max}



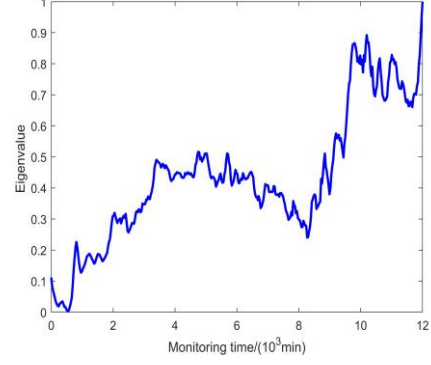
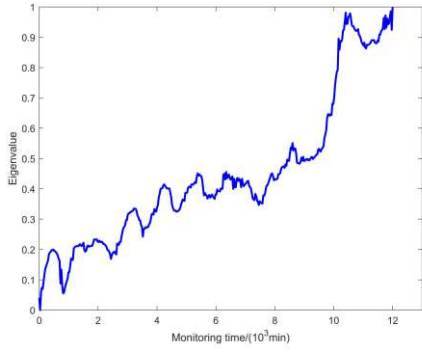
(f) Normalization of x_{If}



(g) Normalization of x_{clf}



(h) Normalization of x_{cf}



(i) Maximum feature of a_2 temperature data (j) Normalization of a_3 driven torque data

Fig. 6. Normalisation of vibration data, temperature data, and torque data features

According to Table 5, the RMS index exhibits the strongest correlation with the torque signal. Accordingly, the RMS feature is retained for torque, whereas the maximum-value feature is selected for temperature (Fig. 6i). To eliminate the influence of differences in sensor data dimensions and magnitudes, the extracted features of vibration, temperature, and torque data are normalised (Fig. 6).

Figure 6 shows the variation in the RMS characteristics of the vibration and torque data. Although the RMS characteristics exhibit local fluctuations, they display an overall upward trend as monitoring progresses. This trend is mainly attributed to wear and fatigue accumulation in the equipment during continuous operation. In addition, owing to the nonlinear and non-monotonic characteristics of wind-turbine slewing bearings, the RMS feature more effectively reflects changes in the operating condition of the bearing. Therefore, the RMS feature is selected as the sensitive feature of the original torque data, as it effectively characterises the operating state of the wind-turbine slewing bearing. Furthermore, the selected feature is computationally efficient, making it well suited to practical data analysis.

According to the analysis, the eight selected features extracted from the vibration data, along with the maximum-value characteristics of temperature data and the RMS characteristics of torque data, are integrated for model validation and RUL prediction.

5.1 State Estimation and Degradation Model Validation

The proposed filter is assessed using the ESS defined in Eq. (12).

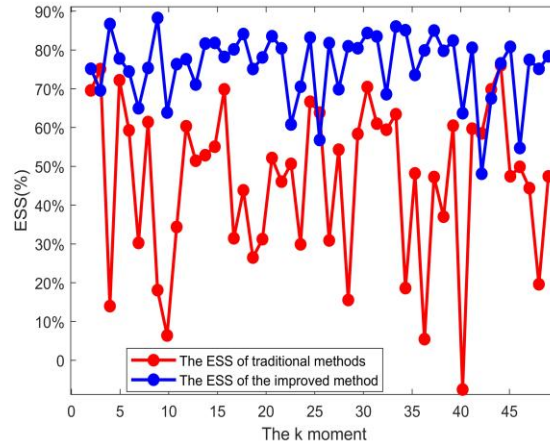


Fig. 7. Comparison of ESS values for two APF posterior PDFs

During the first 50 monitoring instants, the enhanced APF consistently achieves higher ESS values than the conventional APF at most time points (Fig. 7), indicating reduced weight concentration and a larger effective particle population.

The eight features extracted from the vibration data, along with the maximum-value feature of the temperature data and the maximum-value feature of the torque data, are integrated to estimate the degradation states and model parameters for RUL prediction. In the literature^[33], the failure threshold is determined through extensive destructive testing. The degradation data of wind-turbine slewing bearings of the same type are then normalised, and the average failure threshold is defined as 1. For ease of analysis, the estimated degradation states are denoted by x_1 , x_2 , x_3 , x_4 , x_5 , x_6 , x_7 , x_8 , x_9 , and x_{10} , respectively. The 10 estimated degradation states are subsequently fused to obtain the integrated degradation state x_{11} , resulting in a total of 11 degradation states (Fig. 8).

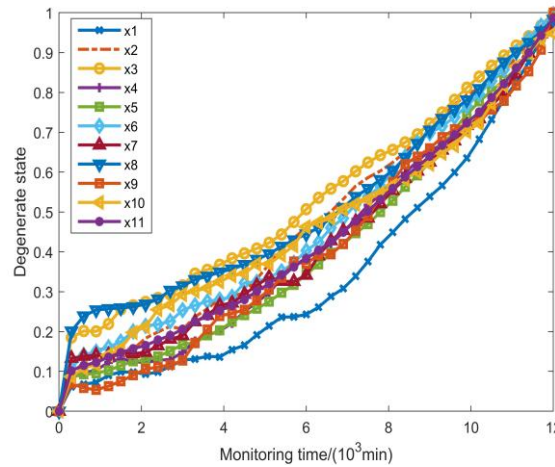


Fig. 8. Estimated degradation state curve

The degradation-state trajectories (Fig. 8) exhibit distinct evolution rates and local fluctuations. These differences indicate that individual features respond differently to random operating conditions, whereas the fused trajectory captures the overall degradation trend.

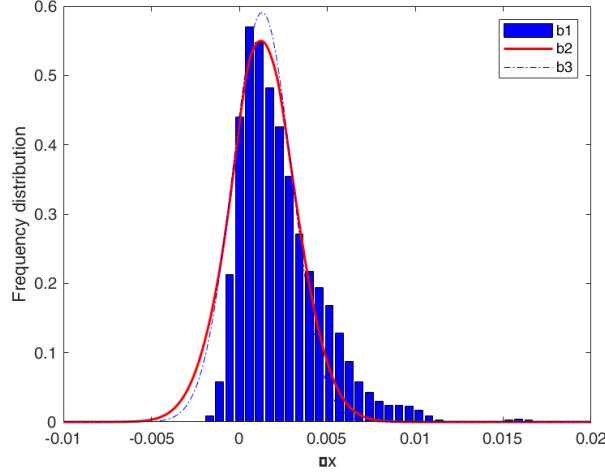


Fig. 9. Fitting results for the Δx_{11} degradation-state increment

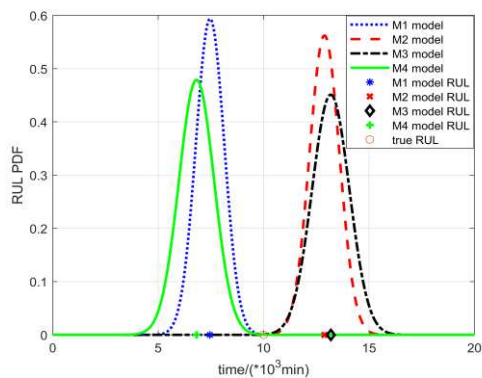
To verify the applicability of Eq. (8), a Monte Carlo simulation is conducted to obtain the PDF curve of the degradation state increment Δx_i (Fig. 9). First, the increments Δx_i between adjacent degradation states are calculated, and their frequencies are utilised to construct a histogram, denoted by b_1 . Second, a fitted curve is obtained from the histogram and is denoted by b_2 . Finally, the mean and variance of the degradation state increment are calculated to derive a Gaussian PDF, denoted by b_3 .

Figure 9 presents the frequency distribution of the degradation-state increment Δx_{11} obtained from the Monte Carlo simulation. The fitted curve b_2 , derived from the histogram of the degradation-state increments, closely matches the Gaussian PDF curve b_3 , indicating that the degradation-state increment Δx_{11} follows a Gaussian distribution, thereby supporting the applicability of Eq. (8) for modelling the degradation process of the system.

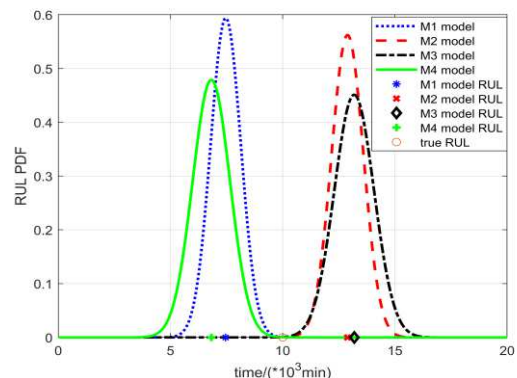
5.2 RUL Prediction

To verify the applicability of the proposed RUL prediction method, several representative life prediction methods are selected for comparison. The method proposed in this study, denoted by model M_1 , incorporates both the effects of random stress induced by external loading and the interactions between different features and

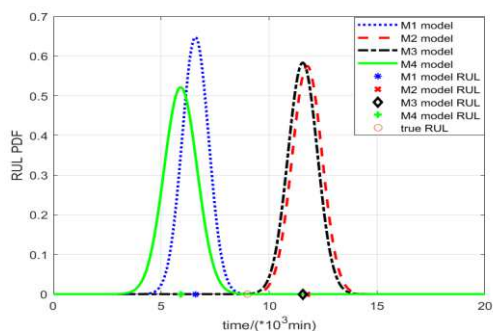
degradation states. A reduced model that excludes the effects of random stress is denoted by M_2 . A further reduced model that excludes both the interactions between different features and degradation states and the effects of random stress is denoted by M_3 ^[30]. A nonlinear Wiener process model incorporating multiple fused features for RUL prediction is denoted by M_4 ^[19,34,35].



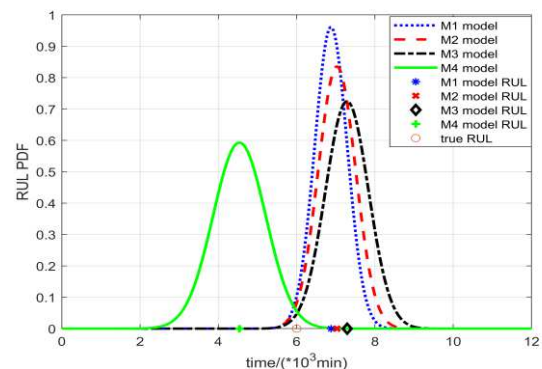
(a) $t_k = 1 \times 10^3$ min RUL prediction



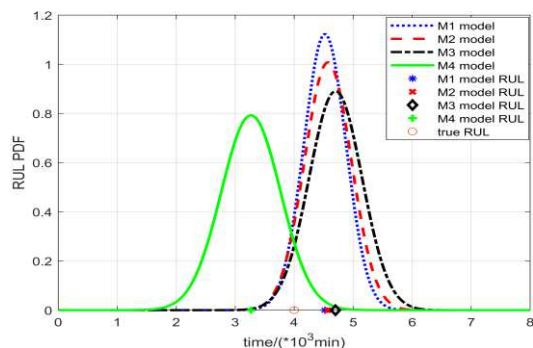
(b) $t_k = 2 \times 10^3$ min RUL prediction



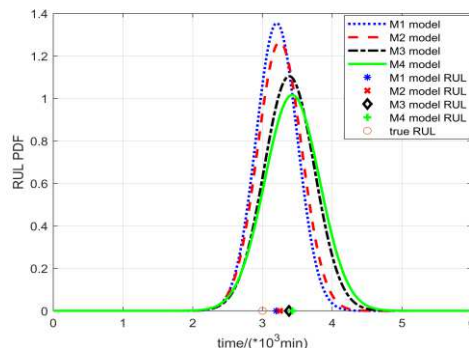
(c) $t_k = 3 \times 10^3$ min RUL prediction



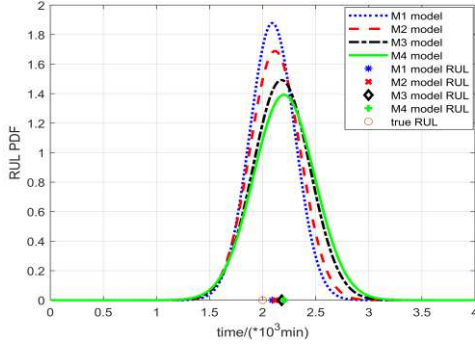
(d) $t_k = 6 \times 10^3$ min RUL prediction



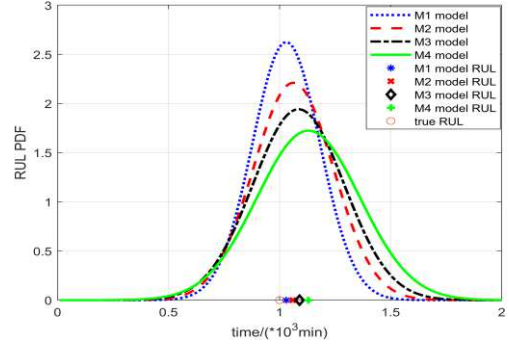
(e) $t_k = 8 \times 10^3$ min RUL prediction



(f) $t_k = 9 \times 10^3$ min RUL prediction



(g) $t_k = 10 \times 10^3$ min RUL prediction

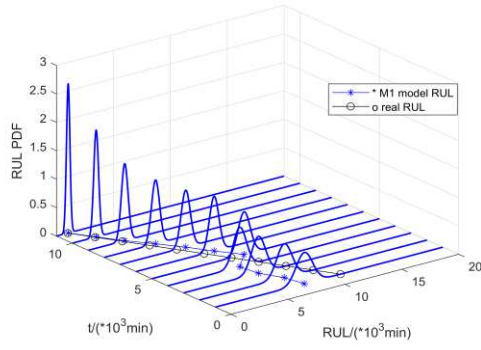


(h) $t_k = 11 \times 10^3$ min RUL prediction

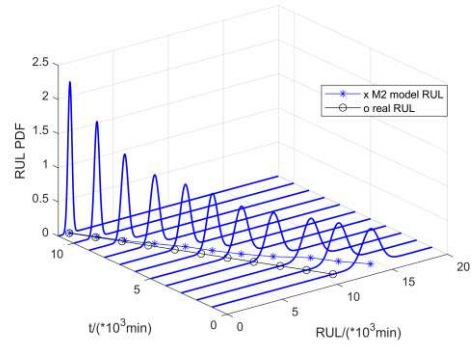
Fig. 10. Comparison of RUL prediction results at different monitoring times.

To evaluate the effectiveness of the proposed model, the RUL prediction results of the M_1 , M_2 , M_3 , and M_4 models are compared at different monitoring times (Figs. 10 and 11).

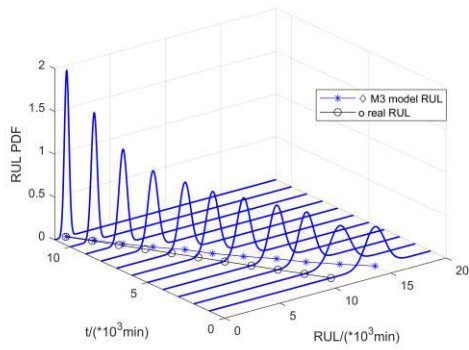
Notably, with the passage of time and the continuous increase of monitoring data, the prediction errors of all four models gradually decrease, indicating that the model parameters are progressively refined (Fig. 10). The prediction accuracy of the model improves significantly, and the actual RUL of the wind-turbine slewing bearing is generally encompassed within the predicted PDF intervals throughout the observation period. Overall, the M_1 model exhibits the smallest prediction error, with its estimated RUL remaining closest to the actual RUL. This superior performance can be attributed to the incorporation of both random-stress effects and the interactions between different features and degradation states, resulting in a degradation model that more closely reflects the actual operating conditions of the slewing bearing and consequently achieves higher RUL prediction accuracy. In contrast, the M_2 model employs the same degradation modelling framework but excludes the effects of random stress, leading to reduced RUL prediction accuracy. The M_3 model does not account for the interactions between different features and degradation states or the effects of environmental factors on the operating state of the slewing bearing. Consequently, its RUL prediction accuracy is lower than that of M_1 and M_2 . Similarly, the model employs a centralised fusion approach to establish the degradation process for RUL prediction but does not incorporate the effects of environmental factors. In addition, it ignores the interactions between degradation states and the random correlations among different features. Compared with the other models, this model exhibits the lowest RUL prediction accuracy. Overall, these results demonstrate that the proposed framework improves the accuracy of RUL prediction for wind-turbine slewing bearings.



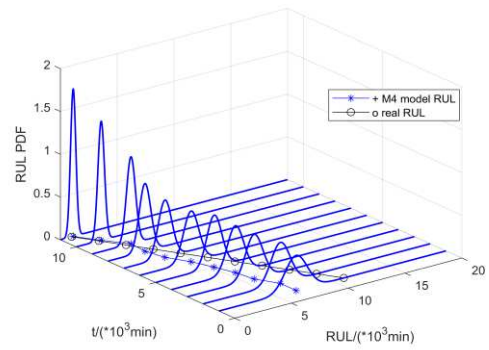
(a) RUL prediction of the M_1 model



(b) RUL prediction of the M_2 model



(c) RUL prediction of the M_3 model



(d) RUL prediction of the M_4 model

Fig. 11. Comparison of the RUL PDFs of the four models.

Notably, with the increase of monitoring time, the RUL prediction PDF curve of the M_1 model becomes narrower and exhibits a higher amplitude (Fig. 11). This finding indicates that as additional monitoring data become available, the model parameters are continuously updated, thereby reducing the uncertainty of the RUL prediction and producing estimates that are progressively closer to the actual value. Overall, the M_{s1} model outperforms the other three models in terms of RUL prediction accuracy. These results demonstrate that the proposed framework effectively captures the degradation process of the wind-turbine slewing bearing and provides more accurate RUL predictions.

To further evaluate the RUL prediction performance of the four models, four performance metrics, namely, MSE , SOA , A , and $PICP$, were calculated^[36-38]. Table 6 and Figures 12–14 present the comparative results.

Table 6

Error metrics calculation formula (l denotes the actual value and $\tilde{l}$ denotes the predicted value).

Error index	Formula	Error index	Formula
Mean square error(MSE)	$MSE = \frac{1}{n} \sum_{i=1}^n (l_i - \tilde{l}_i)^2$	Score of accuracy (SOA)	$SOA_i = \begin{cases} \exp\left(-\ln(0.5)\left(100 \times \frac{l_i - \tilde{l}_i}{l_i} / 5\right)\right), & l_i - \tilde{l}_i < 0 \\ \exp\left(\ln(0.5)\left(100 \times \frac{l_i - \tilde{l}_i}{l_i} / 20\right)\right), & l_i - \tilde{l}_i > 0 \end{cases}$
Prediction interval coverage probability(PICP)	$PICP = 100\% \cdot \frac{1}{N} \sum_{i=1}^N I(l_i)$ $I(l_i) = \begin{cases} 1, D_\alpha(l'_i) < l_i < U_\alpha(l'_i) \\ 0, & else \end{cases}$	Accuracy (A)	$A = \frac{100\%}{N} \sum_{i=1}^N g(l_i - l'_i)$ $g(l_i - l'_i) = \begin{cases} 1, -2.04 < l_i - l'_i < 2.45 \\ 0, & else \end{cases}$

The larger the SOA value, the more accurate the RUL estimation. In addition, the smallest MSE indicates the lowest uncertainty of RUL estimation. Although SOA_i exhibits some fluctuations, at almost all measurement points, the RUL prediction of the M1 model is more accurate compared with that of the M_2 , M_3 , and M_4 models (Fig. 12). Furthermore, as the monitoring time increases and additional monitoring data become available, the model parameters are progressively updated, resulting in a gradual reduction in RUL prediction uncertainty (Fig. 13). Compared with that of the other three models, the MSE_i of the M_1 model remains consistently lower, indicating superior predictive performance. Figure 14 shows the overall SOA and MSE of the four models, and Table 7 presents the corresponding RUL prediction errors.

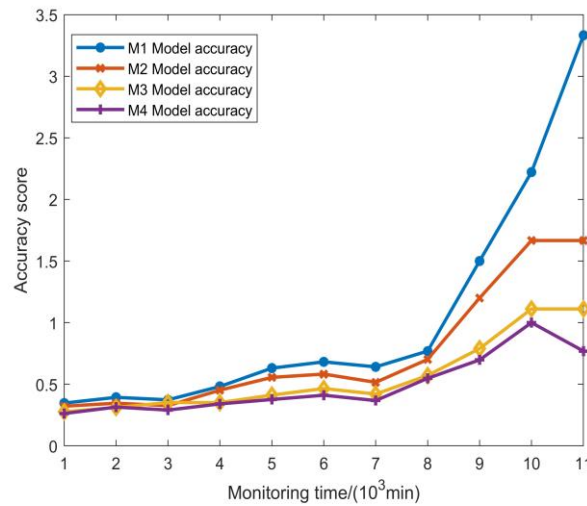


Fig. 12. Comparison of the four models in terms of SOA_i for RUL prediction

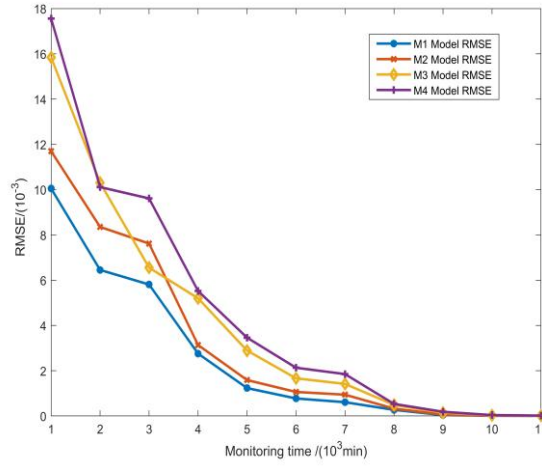


Fig. 13. Comparison of the four models in terms of MSE_i for RUL prediction

Notably, the values of SOA decrease in the following order: $M_1 < M_2 < M_3 < M_4$ (Fig. 14). Therefore, the M_1 model achieves more accurate RUL predictions compared with the other three models. Similarly, the M_1 model achieves the smallest MSE compared with the other three models. Therefore, it produces the smallest RUL prediction error and the highest prediction accuracy.

Notably, compared with the other three models, the M_1 model achieves the best performance in A (85.50%) and $PICP$ (96.52%) (Table 7). These results further demonstrate that the M_1 model provides more accurate RUL prediction compared with the other models.

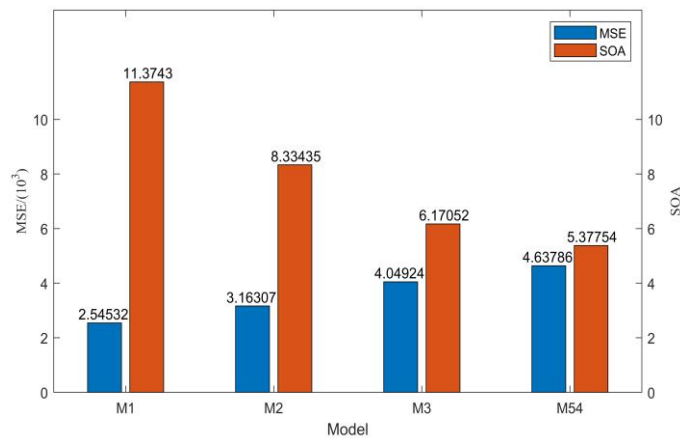


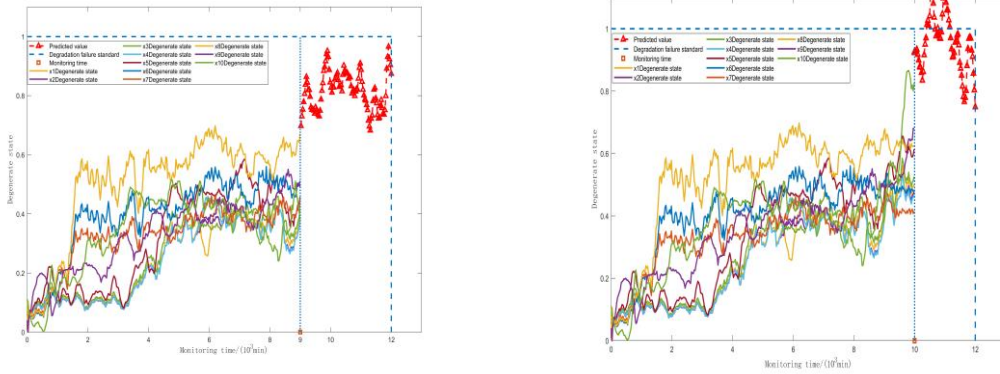
Fig. 14. Comparison of the four models in terms of SOA and MSE for RUL prediction

Table 7. Comparison of the prediction accuracy and uncertainty quantification of the four methods

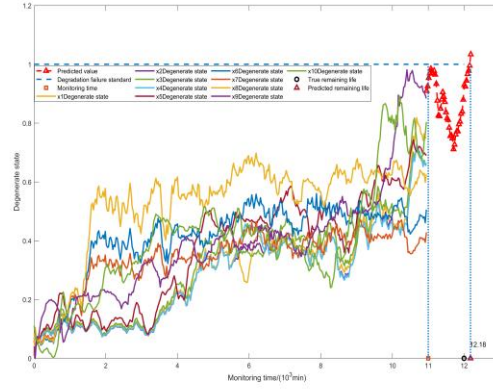
Models	Method			
	SOA	MSE (10^3)	A (%)	$PICP$ (%)
M_1	11.37	2.55	85.50	96.52
M_2	8.33	3.16	81.25	95.91
M_3	6.17	4.05	75.00	87.14
M_4	5.38	4.64	63.00	80.35

Models	Method			
	SOA	MSE (10^3)	A (%)	$PICP$ (%)
M_1	11.37	2.55	85.50	96.52
M_2	8.33	3.16	81.25	95.91
M_3	6.17	4.05	75.00	87.14
M_4	5.38	4.64	63.00	80.35

For an additional data-driven benchmark, a three-layer CNN is trained using the same 10 input features, and its predictions are compared with those of M_1 at three different monitoring times (Fig. 15).



(a) $t_k = 9 \times 10^3$ min degradation state prediction (b) $t_k = 10 \times 10^3$ min degradation state prediction



(c) $t_k = 11 \times 10^3$ min degradation state prediction

Fig. 15. Degradation state prediction results of the CNN model at different monitoring times

As the training time and training data increase, the prediction error of the CNN model decreases progressively (Fig. 15). At a monitoring time of 11×10^3 min, the average failure time predicted by the CNN model is 12.18×10^3 min, while the corresponding prediction by the M_1 model is 12.03×10^3 min (Fig. 15c), which is closer to the actual average failure time of the slewing bearing, indicating superior prediction accuracy. These results indicate that at the same monitoring point, the M_1 model achieves higher RUL prediction accuracy than the CNN model.

Collectively, the ablation study and CNN comparison demonstrate that random-stress modelling, feature-state interactions, and the enhanced APF algorithm each contribute to the improved prediction performance. Compared with the three reduced stochastic models and the CNN benchmark, the complete framework achieves the highest RUL prediction accuracy in the reported experiments.

6 Conclusion

To address the accelerated degradation of wind-turbine slewing bearings caused by random load variations and the influence of multiple features on different degradation states, a random-stress model is incorporated into the state transfer equation, and a nonlinear state-space model that captures the interactions between different features and degradation states is established. In addition, to address the problems of sample degradation and scarcity in the conventional APF, an adaptive enhanced APF algorithm is employed to improve the accuracy of state estimation. Bayesian hyperparameter estimation is used to model the distribution of randomly varying stress, while distributed state fusion is achieved using a diagonal weighting

matrix. Finally, the proposed prediction model is validated using vibration, temperature, and torque monitoring data. The results demonstrate that incorporating the effects of randomly varying stress and the interactions between different features and degradation states effectively improves the RUL prediction accuracy for wind-turbine slewing bearings. These findings indicate that the proposed framework is well suited for real-time RUL prediction for wind-turbine slewing bearings. The results also reveal several limitations of this model; particularly when handling a larger number of features, the model becomes more complex, requires greater computational resources, and adversely affects the timeliness of predictions.

References:

- [1] Ding P, Wang H, Bao W, et al. HYGP-MSAM based model for slewing bearing residual useful life prediction[J]. *Measurement*, 2019, 141: 162-175.
- [2] Liu J, Hou B, Lu M, et al. Box-Cox transformation based state-space modeling as a unified prognostic framework for degradation linearization and RUL prediction enhancement[J]. *Reliability Engineering & System Safety*, 2024, 244: 109952.
- [3] Zio E. Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice[J]. *Reliability Engineering & System Safety*, 2022, 218: 108119.
- [4] Qin Y, Yang J, Zhou J, et al. A new supervised multi-head self-attention autoencoder for health indicator construction and similarity-based machinery RUL prediction[J]. *Advanced Engineering Informatics*, 2023, 56: 101973.
- [5] Wang Y, Li Y, Zhang Y, et al. Incorporating prior knowledge into self-supervised representation learning for long PHM signal[J]. *Reliability Engineering & System Safety*, 2024, 241: 109602.
- [6] Xu X, Zhou J, Weng X, et al. A Novel Evidence Reasoning-Based RUL Prediction Method Integrating Uncertainty Information[J]. *Reliability Engineering & System Safety*, 2024: 110250.
- [7] Lundberg G, Palmgren A. Dynamic capacity of rolling bearings[J], 1949.
- [8] Ioannides E, Harris T A. A new fatigue life model for rolling bearings[J], 1985.
- [9] Zaretsky E V, Poplawski J V, Miller C. Rolling bearing life prediction-past, present, and future[C]. *International Tribology Conference 2000*, 2000.
- [10] Zhou R, Cheng H S, Mura T. Micropitting in rolling and sliding contact under mixed lubrication[J], 1989.

- [11] Zhou R S. Surface topography and fatigue life of rolling contact bearing[J]. Tribology Transactions, 1993, 36(3): 329-340.
- [12] Ansari S, Ayob A, Lipu M H, et al. Remaining useful life prediction for lithium-ion battery storage system: A comprehensive review of methods, key factors, issues and future outlook[J]. Energy Reports, 2022, 8: 12153-12185.
- [13] Li X, Ding Q, Sun J-Q. Remaining useful life estimation in prognostics using deep convolution neural networks[J]. Reliability Engineering & System Safety, 2018, 172: 1-11.
- [14] Liu R, Yang B, Hauptmann A G. Simultaneous bearing fault recognition and remaining useful life prediction using joint-loss convolutional neural network[J]. IEEE Transactions on industrial informatics, 2019, 16(1): 87-96.
- [15] Yang B, Liu R, Zio E. Remaining useful life prediction based on a double-convolutional neural network architecture[J]. IEEE Transactions on Industrial Electronics, 2019, 66(12): 9521-9530.
- [16] Zhang X, Xiao L, Kang J. Degradation prediction model based on a neural network with dynamic windows[J]. Sensors, 2015, 15(3): 6996-7015.
- [17] Ye Z S, Xie M. Stochastic modelling and analysis of degradation for highly reliable products[J]. Applied Stochastic Models in Business and Industry, 2015, 31(1): 16-32.
- [18] Xiao-Sheng S, Chang-Hua H, Dong-Hua Z. Nonlinear degradation process modeling and remaining useful life estimation subject to measurement error[J]. Acta Automatica Sinica, 2014, 39(5): 530-541.
- [19] Zhang Z, Si X, Hu C, et al. Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods[J]. European Journal of Operational Research, 2018, 271(3): 775-796.
- [20] Si X-S, Wang W, Hu C-H, et al. A Wiener-process-based degradation model with a recursive filter algorithm for remaining useful life estimation[J]. Mechanical Systems and Signal Processing, 2013, 35(1-2): 219-237.
- [21] Si X-S, Wang W, Hu C-H, et al. Remaining useful life estimation based on a nonlinear diffusion degradation process[J]. IEEE Transactions on reliability, 2012, 61(1): 50-67.
- [22] Li W, Jiao Z, Du L, et al. An indirect RUL prognosis for lithium-ion battery under vibration stress using Elman neural network[J]. International Journal of Hydrogen Energy, 2019, 44(23): 12270-12276.

[23] Duan X, Shan J, Xie C, et al. Reliability modeling of pre-stressed composite structures subject to interlayer slip failure under multi-axial random impact and vibration loads[J]. Probabilistic Engineering Mechanics, 2024: 103650.

[24] Li Y, Gao H, Chen H, et al. Accelerated degradation testing for lifetime analysis considering random effects and the influence of stress and measurement errors[J]. Reliability Engineering & System Safety, 2024, 247: 110101.

[25] Wang J-J, Shang D-G, Li W-L, et al. Fatigue life prediction considering conversion of mean stress for titanium alloy under multiaxial thermo-mechanical random loading[J]. International Journal of Fatigue, 2024, 188: 108488.

[26] Xi X, Chen M, Zhou D. Remaining useful life prediction for multi-component systems with hidden dependencies[J]. Science China Information Sciences, 2019, 62: 1-16.

[27] Zhang X, Liu D, Lei B, et al. An intelligent particle filter with resampling of multi-population cooperation[J]. Digital Signal Processing, 2021, 115: 103084.

[28] An D, Choi J-H, Kim N H. Prognostics 101: A tutorial for particle filter-based prognostics algorithm using Matlab[J]. Reliability Engineering & System Safety, 2013, 115: 161-169.

[29] Si X-S, Wang W, Chen M-Y, et al. A degradation path-dependent approach for remaining useful life estimation with an exact and closed-form solution[J]. European Journal of Operational Research, 2013, 226(1): 53-66.

[30] Feng Y, Huang X, Hong R, et al. A multi-dimensional data-driven method for large-size slewing bearings performance degradation assessment[J]. Journal of Central South University (Science and Technology), 2017, 48(03): 684-693.

[31] He P, Hong R, Wang H, et al. Fatigue life analysis of slewing bearings in wind turbines[J]. International Journal of Fatigue, 2018, 111: 233-242.

[32] Liu M, Dong Z, Shi H. Multi-Sensor Information Fusion and Multi-Model Fusion-Based Remaining Useful Life Prediction of Fan Slewing Bearings with the Nonlinear Wiener Process[J]. Sustainability, 2023, 15(15): 12010.

[33] Tian S, Wang H, Hong R. Residual life assessment of slewing bearing based on multivariate eigenvalues fusion and support vector regression[J]. Journal of nanjing tech university (natural science edition), 2016, 38(3): 50-57.

[34] Fang X, Paynabar K, Gebraeel N. Multistream sensor fusion-based prognostics model for systems with single failure modes[J]. Reliability Engineering & System Safety, 2017, 159: 322-331.

[35] Tse P W, Wang D. Enhancing the abilities in assessing slurry pumps' performance degradation and estimating their remaining useful lives by using captured vibration signals[J]. Journal of Vibration and Control, 2017, 23(12): 1925-1937.

[36] Pang Z, Si X, Hu C, et al. A Bayesian inference for remaining useful life estimation by fusing accelerated degradation data and condition monitoring data[J]. Reliability Engineering & System Safety, 2021, 208: 107341.

[37] Nguyen K T, Medjaher K, Gogu C. Probabilistic deep learning methodology for uncertainty quantification of remaining useful lifetime of multi-component systems[J]. Reliability Engineering & System Safety, 2022, 222: 108383.

[38] Cho K S, Tony Ng H K. Tolerance intervals in statistical software and robustness under model misspecification[J]. Journal of Statistical Distributions and Applications, 2021, 8(1): 10.

CRedit authorship contribution statement

Mingjun Liu: Writing-review & editing, Writing-original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Zengshou Dong: Writing-review & editing, Writing -original draft, Visualization, Supervision, Software, Project administration, Methodology, Investigation. Zhensen Gao: Writing-review, Writing -original draft, Visualization, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Data availability: Data will be made available on request.

Acknowledgements

This work was supported by the Natural Science Foundation of Shanxi Province (202203021211205), the Xinzhou Science and Technology Bureau Basic Research Project (20230505), and the 2025 University-Level Scientific Research Cultivation Program of Xinzhou Normal University (X20250096).