ORIGINAL ARTICLE

Health condition evaluation method for motorized spindle on the basis of optimised VMD and GMM-HMM

Haiji Yang[1,2] · Guofa Li[1,2] · Jialong He[1,2] · Liding Wang[1,2] · Xinyu Nie[3]

Abstract

When the motorized spindles fail, the vibration signal at the bearings will contain the information related to the fault degree and operation condition. Extracting and utilising this information in an industrial environment with strong noise are the key issues of the health condition evaluation of the motorized spindles. In this paper, a health condition evaluation method for motorized spindles on the basis of optimised variational mode decomposition (VMD) and Gaussian mixture model-hidden markov model (GMM-HMM) is proposed. Firstly, using the composite index *KEI* as the fitness function, the parameters in the VMD are optimised by the sooty tern optimisation algorithm (STOA), and a low-dimensional feature matrix that can represent the health condition of a motorized spindle is further constructed. Secondly, the GMM-HMM of each health condition is trained, and the health condition of the motorized spindle is evaluated based on the library. Finally, the hybrid simulation signal is analysed to verify the effectiveness and superiority of the optimised VMD. The rotor unbalanced fault experiment is carried out by using the motorized spindle performance monitoring test platform. The proposed method is used to evaluate the health of the tested motorized spindle, and the results verify its superiority.

Keywords Motorized spindle · Health condition evaluation · Optimised VMD · Composite index *KEI* · GMM-HMM

1 Introduction

As a key functional component of CNC machine tools, CNC machine tool manufacturers have focused on motorized spindle for its advantages of compact structure, low noise and fast response [1]. The performance degradation of motorized spindle seriously affects the machining accuracy and efficiency of CNC machine tools and causes serious economic losses [2]. Using the monitoring information collected by sensors to infer the health condition of the motorized spindle and further develop maintenance strategies is an effective means of preventing motorized spindle failure [3, 4].

✉ Guofa Li
  ligf@jlu.edu.cn

✉ Jialong He
  hejl@jlu.edu.cn

1 Present Address: Key Laboratory of CNC Equipment Reliability, Ministry of Education, Jilin University, Changchun 130025, China

2 School of Mechanical and Aerospace Engineering, Jilin University, Changchun 130025, China

3 Faw Jiefang Automot CO., LTD, Changchun 130025, China

When the motorized spindle fails, the vibration signal at the bearings shows nonlinear and non-stationary characteristics. For such signals, the commonly used signal pre-processing methods include wavelet packet transform (WPT), empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), and local mean decomposition (LMD) [5–8]. The combination of the above methods with machine learning to obtain accurate results has aroused the research interest of scholars. Fei proposed the combination of WPT and SVM to realise the fault evaluation of rolling bearings [9]. Sun et al. proposed that EEMD to decompose vibration signals and used PSO-SVM to identify the faults of bevel gears [10]. Tian et al. proposed to use of LMD and SVD to process vibration signals and then ELM to identify rolling bearing faults [11]. Although the above methods have been applied, some shortcomings remain; for example, WPT requires the artificial selection of parameters [12]. EMD has some problems, such as modal aliasing and the endpoint effect. Although LMD can reduce the mode aliasing to a certain extent, it is greatly affected by noise [13]. EEMD is limited by its large amount of computation and requires human experience to determine the added white noise amplitude and the number of iterations [14]. VMD is an adaptive,

non-recursive modal variational signal processing method that can overcome the shortcomings of the above methods to a certain extent [15, 16]. An and Zhang proposed to use of VMD to decompose the vibration signals of bearing loosening and gear faults and carry out the envelope demodulation analysis of sensitive modal components [17].

However, the traditional VMD is susceptible to the influence of penalty factor α and decomposition layers K, and the setting of these two key parameters is a problem that must be urgently solved. Kumar et al. used kernel mutual information as a fitness function, genetic algorithm to optimise two parameters, and the optimised VMD to extract the early fault features of bearings [18]. Zhong et al. proposed to use the whale algorithm to optimize two parameters, and proposed a millimeter-wave radar FOD detection method based on the optimised VMD [19]. To sum up, on the basis of a reasonable fitness function, the penalty factor α and decomposition layers K in the VMD can be optimised by using a heuristic algorithm. In addition, the performance of the heuristic algorithm can determine the computational efficiency of this process. Therefore, the STOA is proposed to optimise two parameters in the VMD. In terms of convergence and computational complexity, the STOA is highly competitive with nine other heuristic algorithms [20].

The essence of data-driven health condition evaluation is to establish the mapping relationship between the feature space and the health condition. Compared with some intelligent algorithms such as artificial neural networks (ANN), decision tree (DT), support vector machine (SVM), hidden Markov model (HMM) has the advantages of simple structural expression and strong physical interpretation, which can be better applied to the health condition evaluation of motorized spindle [21–23]. However, the signal measured by a sensor is a continuous signal. Although this signal can be discretised, the use of limited discrete symbols to represent continuous observation variables will inevitably lead to information loss. Aiming at the above problems, the continuously fault features can be represented by introducing a probability density function. Theoretically, the Gaussian mixture model (GMM) can use several weighted Gaussian distributions to approximate the probability density function of the observation matrix in each condition. GMM-HMM has better modelling and analysis abilities than HMM. Liu et al. proposed to use of GMM-HMM for bearing fault diagnosis, input the test data into the trained GMM-HMM, and output the bearing state [24]. Zheng and Gao proposed to use GMM-HMM to create classifiers for diagnosing the downdole operating conditions of sucker rod pumping [25].

This paper uses the optimised VMD to remove the noise component in the original signal and combines the dynamic modelling ability of GMM-HMM to construct the health condition evaluation model of the motorized spindle. The rest of this paper is arranged as follows. In Sect. 2, the related methods of this paper are briefly introduced. In Sect. 3, the steps of the proposed method are described in detail. In Sect. 4, the hybrid simulation signal is analysed by using the optimised VMD to verify the superiority of this method. In Sect. 5, the rotor unbalance fault experiment is carried out on a motorized spindle, and the proposed method is used to diagnose the health condition to verify the effectiveness and superiority of the proposed method. The paper is summarised in Sect. 6.

2 Methodology

2.1 Optimised VMD

VMD has a complete mathematical theoretical basis and can decompose the original signal (f) into a series of intrinsic mode functions (IMFs) nonlinearly and determine the frequency centre and bandwidth of the IMFs. The constrained variational model can be expressed as follows:

$$\begin{cases} \min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] \cdot e^{-j\omega_k t} \right\|_2^2 \right\} \\ s.t \quad \sum_k u_k = f \end{cases} \quad (1)$$

where $u_k(t)$ is the intrinsic modal function, ω_k is the centre frequency, ∂_t is the gradient relative to the time series, and $\delta(t)$ is the unit pulse function.

Equation (1) is transformed into an unconstrained variational problem using penalty factor α and Lagrange multiplier $\lambda(t)$ as follows:

$$L\left(u_k, \omega_k, \lambda\right) = \alpha \sum_k \left\| \partial_t[(\delta(t) + \frac{j}{\pi t}) * u_k(t)] \cdot e^{-j\omega_k t} \right\|_2^2 \\ + \left\| f - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f - \sum_k u_k(t) \right\rangle \quad (2)$$

An alternating direction multiplier algorithm is used to update *relevant parameters* as follows:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \hat{\lambda}(\omega)/2}{1 + 2\alpha(\omega - \omega_k)^2} \quad (3)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \quad (4)$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau[\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega)] \quad (5)$$

where τ is the fidelity coefficient, ^ represents the Fourier transform, and n is the number of iterations Given discrimination

accuracy ε, i.e. $\varepsilon > 0$, if Eq. (6) is satisfied, then the iteration is stopped.

$$\sum_k \left\| u_k^{n+1} - u_k^n \right\|_2^2 \Big/ \left\| u_k^n \right\|_2^2 < \varepsilon \tag{6}$$

However, the traditional VMD is susceptible to the influence of penalty factor α and decomposition layers K. When penalty factor α is improperly selected, VMD cannot obtain the optimal bandwidth IMF. When decomposition layers K is set improperly, over-decomposition or under-decomposition will occur. To solve this problem, this paper proposes to optimise the two parameters of the VMD by using the STOA and takes composite index KEI as the fitness function. Sooty terns are marine gregarious birds that mainly feed on fish, earthworms and insects. The racial characteristics of sooty terns include migratory and aggressive behaviours, and the mathematical model is as follows:

Migration behaviour: During migration, sooty terns should meet three conditions, namely, avoiding conflict, gathering and updating. The relevant descriptions are shown in Eqs. (7)–(9).

$$\begin{cases} \overline{C}_{st} = S_A \times \overline{P}_{st}(z) \\ S_A = C_f - [z \times (C_f / Max_{iterations})] \\ z = 1, 2, 3 \cdots, Max_{iterations} \end{cases} \tag{7}$$

where $\overline{C}_{st}$ represents the position of an individual that does not collide with other individuals, $\overline{P}_{st}(z)$ represents the current location of the individual, S_A is a variable used to avoid conflicts between adjacent individuals, C_f is the control variable used to adjust S_A, and z represents the iteration times.

$$\begin{cases} \overline{M}_{st} = C_B \times [\overline{P}_{bst}(z) - \overline{P}_{st}(z)] \\ C_B = 0.5 \times R_{and} \end{cases} \tag{8}$$

where $\overline{M}_{st}$ represents the different locations of individual $\overline{P}_{st}$ towards best fit individual $\overline{P}_{bst}(z)$, C_B is the random variable responsible for improved exploration, and R_{and} is a random number between 0 and 1.

$$\overline{D}_{st} = \overline{C}_{st} + \overline{M}_{st} \tag{9}$$

where $\overline{D}_{st}$ represents the gap between the individual and best fit individual.

Aggressive behaviour: When sooty terns attack their prey, their spiral behaviour in the air can be described as follows:

$$\begin{cases} x' = R_{adius} \times \sin(i) \\ y' = R_{adius} \times \cos(i) \\ z' = R_{adius} \times i \\ r = u \times e^{kv} \end{cases} \tag{10}$$

where R_{adius} represents the radius of each spiral, i is a variable between 0 and 2π, and u and v are constants for defining the spiral shape.

According to Eq. (10), the position of the updated individual is as follows:

$$\overline{P}_{st}(z) = \left[\overline{D}_{st} \times (x' + y' + z') \right] \times \overline{P}_{bst}(z) \tag{11}$$

where $\overline{P}_{st}(z)$ represents updating the positions of other search individuals and saving the best optimal solution.

Another key problem of the optimised VMD is the construction of a fitness function. In the field of mechanical fault diagnosis, envelope entropy and kurtosis indexes are mainly used to measure the degree of fault. However, envelope entropy can only reflect the periodicity of the signal but not the impact characteristics of the signal. The kurtosis index only considers the distribution density of the impact signal and ignores the highly dispersed components. In this paper, the characteristics of the two indexes are comprehensively considered, the envelope entropy is modified by the kurtosis index, and fitness function KEI is constructed. The relevant definitions are as follows:

$$KEI = \frac{K_v^k}{H_e^k}, \ k = 1, 2, \cdots K \tag{12}$$

where K_v^k is the kurtosis index of the kth IMF, and H_e^k is the envelope entropy of the kth IMF. The relevant definitions are as follows:

$$\begin{cases} p_k(t_i) = \frac{h_k(t_i)}{\sum_{i=1}^N h_k(t_i)} \\ H_e = \frac{-\sum_{i=1}^N p_k(t_i) \ln p_k(t_i)}{\ln N} \end{cases} \tag{13}$$

where N is the number of data points, and $h_k(t_i)$ is the envelope signal after Hilbert demodulation. H_e is the corresponding envelope entropy.

$$\begin{cases} K_v = \frac{\frac{1}{N-1} \sum_{i=1}^N \left[u_k(t_i) - \overline{u}_k(t) \right]^4}{\sigma_k^4} \\ \overline{u}_k(t) = \frac{1}{N} \sum_{i=1}^N u_k(t_i) \\ \sigma_k = \sqrt{\frac{1}{N-1} \sum_{i=1}^N \left[u_k(t_i) - \overline{u}_k(t) \right]^2} \end{cases} \tag{14}$$

where $\overline{u}_k(t)$ is the mean value of the IMFs, and σ_k is the standard deviation of the IMFs.

Considering that the optimised VMD takes the minimum fitness value as the optimisation objective, the optimisation objective is expressed as follows:

$$\begin{cases} fit = \min \{-KEI\} \\ s.t. \, k_{\min} \leq k \leq k_{\max}, \ \alpha_{\min} \leq \alpha \leq \alpha_{\max} \end{cases} \tag{15}$$

2.2 GMM-HMM

The HMM is a double stochastic process, including a Markov process that describes the transition between conditions and a stochastic process that describes the observed sequence in each condition. The relationship between these processes is shown in Fig. 1. Generally, the HMM consists of the following five elements:

1. Q is the number of hidden states. If the state at time t is denoted as q_t, then $q_t \in \{S_1, S_2, ..., S_Q\}$.
2. M is the number of the observation symbols. If the observations at t time is o_t, then $o_t \in \{v_1, v_2, ..., v_M\}$.
3. A is the state transition probability matrix. $A = \{a_{ij}\}_{Q \times Q}$, and a_{ij} represents the probability of transition from state S_i at time t to state S_j at time $t+1$.

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i), i \geq 1, j \leq Q \quad (16)$$

4. B is the observation probability matrix. $B = \{b_i(v_m)\}_{Q \times M}$, and $b_i(v_m)$ represents the probability that the observation symbol is v_k when the state is S_i at time t.

$$b_i(v_k) = P(v_k | q_t = S_i), 1 \leq k \leq M \quad (17)$$

5. π is the initial state distribution. π_i represents the probability of being in state S_i at the initial time.

$$\pi_i = P(q_1 = S_i), 1 \leq i \leq Q, \sum_{i=1}^{Q} \pi_i = 1 \quad (18)$$

A complete HMM can be represented by $\lambda = (\pi, A, B)$. However, in the actual process of condition monitoring, the observation symbols of the signal are usually continuous variations. Although the continuous signal can be discretised, the use of discrete symbols to represent continuous observation variables will inevitably cause information loss. To solve the above problems, the GMM can be used to approximate the probability density function of the observation matrix in each state. The observations at time
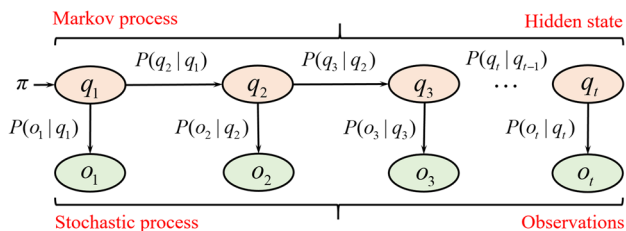
t is assumed to be a D-dimensional vector. For state S_i, the continuous probability distribution function $b_i(o_t)$ can be expressed as follows:

$$\begin{cases} b_i(o_t) = \sum_{m=1}^{M_i} w_{im} b_{im}(o_t) = \sum_{m=1}^{M_i} w_{im} N(o_t, \mu_{im}, Cov_{im}) \\ \sum_{m=1}^{M_i} w_{im} = 1, w_{im} \geq 0 \end{cases} \quad (19)$$

where M_i is the number of Gaussian components, $b_{im}(o_t)$ is the probability density function of the mth component, w_{im} is the weight of the mth Gaussian component, μ_{im} is the mean vector, Cov_{im} is the covariance matrix, and $N(o_t, \mu_{im}, Cov_{im})$ is the Gaussian probability density function:

$$N(o_t, \mu_{im}, Cov_{im}) = \frac{1}{\sqrt{(2\pi)^D |Cov_{im}|}} \exp[-\frac{1}{2}(o_t - \mu_{im})^T Cov_{im}^{-1}(O_t - \mu_{im})] \quad (20)$$

In this part, the algorithms used are the Forward–Backward and Baum–Welch algorithms.

3 Proposed method

In this study, a health condition evaluation method for motorized spindle on the basis of optimised VMD and GMM-HMM is proposed. The flow of the proposed method is shown in Fig. 2.

Step 1 According to the historical information, divide the health condition into L levels and collect the original signals of each level.

Step 2 Initialise the parameters of the STOA.

Step 3 Randomly initialise the search individuals and calculate the fitness value of each search individual. Update the individual location and record the global optimal fitness value. If the iteration condition is reached, then the optimal parameter combination $[\alpha, K]$ is produced. Otherwise, the iteration continues.

Step 4 The original signal is decomposed by the optimised VMD, and the KEI of the IMFs is calculated. The IMFs are sorted by the KEI, and the former $\lceil K/2 \rceil$ IMFs are reconstructed. $\lceil K/2 \rceil$ indicates that $K/2$ is rounded up.

Step 5 Extract the 10-dimensional time (P_1–P_{10}), 5-dimensional frequency (P_{11}–P_{15}) and 10-dimensional scale (P_{16}–P_{25}) domain features of the reconstructed

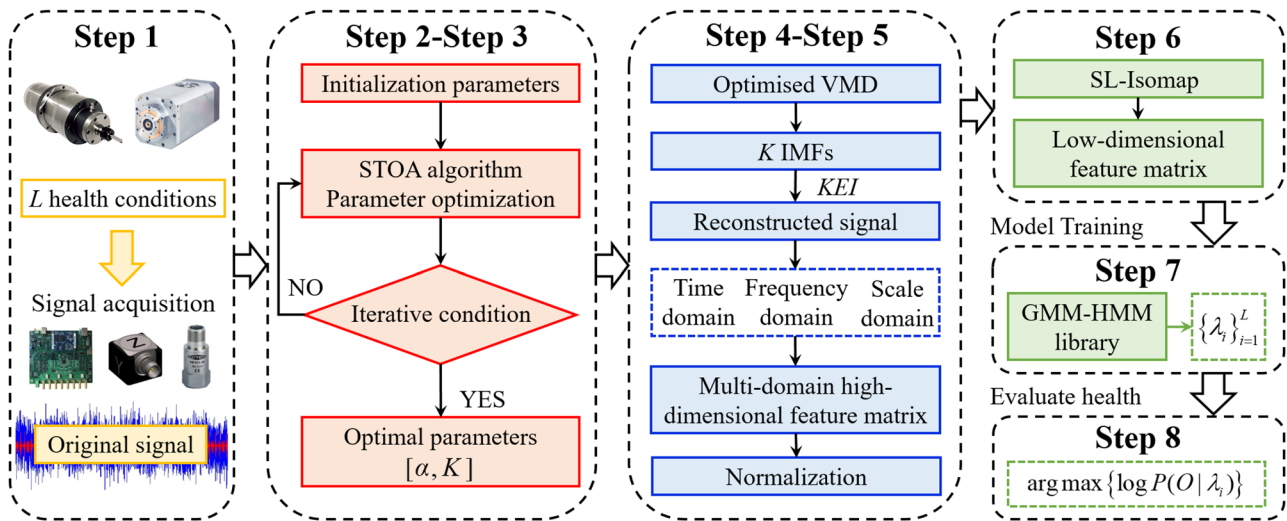**Fig. 1** Schematic of the structure of the HMM

Fig. 2 Flowchart of the proposed method

signals to construct and normalise the multi-domain high-dimensional feature matrix F.

The mathematical expressions of time domain and frequency domain features are shown in Table 1. The main

parameters of multi-scale weighted permutation entropy include embedding dimension $l = 5$, delay time $\tau = 1$ and scale factor $s = 10$ [26, 27].

In Table 1, P1-P10 are time-domain features, and P_{11}-P_{15} are frequency domain features. $x(t)$ is the reconstructed

Table 1 Mathematical expressions of time domain and frequency domain features

Parameter	Expression	Parameter	Expression
Variance	$P_1 = \frac{1}{N} \sum_{i=1}^{N} [x(t) - \bar{x}(t)]^2$	Peak factor	$P_9 = P_5 / P_2$
Root mean square	$P_2 = [\frac{1}{N} \sum_{i=1}^{N} x(t)^2]^{\frac{1}{2}}$	Impulse factor	$P_{10} = P_5 \Big/ \frac{1}{N} \sum_{i=1}^{N} |x(t)|$
Square root amplitude	$P_3 = [\frac{1}{N} \sum_{i=1}^{N} \sqrt{|x(t)|}]^2$	Spectral mean	$P_{11} = \frac{1}{Y} \sum_{y=1}^{Y} s(y)$
Skewness	$P_4 = \sum_{i=1}^{N} [x(t) - \bar{x}(t)]^3 \Big/ (N-1)\sigma^3$	Center of gravity frequency	$P_{12} = \sum_{y=1}^{Y} s(y)f_y \Big/ \sum_{y=1}^{Y} s(y)$
Peak	$P_5 = \max |x(t)|$	Root mean square frequency	$P_{13} = [\sum_{y=1}^{Y} s(y)f_y^2 \Big/ \sum_{y=1}^{Y} s(y)]^{\frac{1}{2}}$
Kurtosis	$P_6 = \sum_{i=1}^{N} [x(t) - \bar{x}(t)]^4 \Big/ (N-1)\sigma^4$	Standard deviation frequency	$P_{14} = [\sum_{y=1}^{Y} (f_y - P_{12})^2 s(y) \Big/ Y - 1]^{\frac{1}{2}}$
Margin factor	$P_7 = P_5 / P_3$	Kurtosis frequency	$P_{15} = \sum_{y=1}^{Y} (f_y - P_{12})^4 s(y) \Big/ (Y-1)P_{14}^4$
Waveform factor	$P_8 = P_2 \Big/ \frac{1}{N} \sum_{i=1}^{N} |x(t)|$		

signal, and N is the number of data points. $s(y)$ is the frequency spectrum of $x(t)$, Y is the number of spectral lines, and f_y is the frequency value of the y-th spectral line. The intermediate parameter σ is defined as follows:

$$\sigma = \sqrt{\sum_{i=1}^{N} [x(t) - \bar{x}(t)]^2 \Big/ N - 1} \tag{21}$$

Step 6 Use the supervised L-Isomap (SL-Isomap) for the secondary feature extraction of multi-domain high-dimensional feature matrix F to obtain low-dimensional feature matrix F' [28, 29].

Step 7 Divide the training and test data according to a certain proportion. Train the GMM-HMM of each health condition using the Baum–Welch algorithm and establish the GMM-HMM library.

Step 8 Substitute the test data into the GMM-HMM library and calculate the log-likelihood probability (LLP) of each model using the Forward–Backward algorithm. Meanwhile, the Softmax function is used to map LLP into [0, 1] to obtain intuitive results, namely:

$$\text{SLLP}_i = = \frac{\exp(\text{LLP}_i)}{\sum_{i=1}^{N} \exp(\text{LLP}_i)}, \ 1 \leq i \leq L \tag{22}$$

The SLLP can reflect the degree of membership between the test data and each model, so the condition corresponding to the maximum SLLP is the most probable health condition, that is,

$$Label(O) = \arg \max\{\text{SLLP}_i\}, \ 1 \leq i \leq L \tag{23}$$

4 Simulation signal analysis

In this section, the hybrid simulation signal $y(t)$ is analysed by using the optimised VMD. The purpose of the analysis is to determine the sensitive IMF with the most fault information

and further extract the fault frequency. The hybrid simulation signal $y(t)$ consists of 4 parts, including fault impact signal $x(t)$, the rotation signal $r(t)$ of other components, random impact signal $h(t)$ and random noise $n(t)$. Among them, fault impact signal $x(t)$ is based on the partial fault of the sun gear in the planetary gearbox, and its mathematical model can be expressed as follows [30]:

$$\begin{cases} x(t) = s(t) \sum_{k=1}^{K} a_k(t) \cos[2\pi k f_m t + b_k(t) + \varphi_k] \\ s(t) = 1 - \cos(2\pi n f_{sr} t) \\ a_k(t) = \sum_{n=1}^{N} A_{kn} \cos(2\pi n f_s t + \alpha_{kn}) \\ b_k(t) = \sum_{n=1}^{N} B_{kn} \cos(2\pi n f_s t + \beta_{kn}) \end{cases} \tag{24}$$

where $a_k(t)$ and $b_k(t)$ are amplitude and frequency modulations, respectively; $s(t)$ is amplitude modulation caused by the rotation of the sun gear; f_m is the gear meshing frequency; f_s is fault frequency; and f_{sr} is the absolute rotation frequency of the sun gear.

The rotation signal $r(t)$ of the other components can be represented by multiple harmonic components as follows:

$$r(t) = \sum_i C_i \sin(2\pi f_i t + \theta_i) \tag{25}$$

where C_i, f_i and θ_i represent the amplitude, frequency and phase of each harmonic component, respectively.

The random impact signal $h(t)$ is shown as follows:

$$\begin{cases} h(t) = \sum_i R_j H(t - Tr_j) \\ H(t) = e^{-\beta t} \sin(2\pi f_{RE} t) \end{cases} \tag{26}$$

where R_j is a random variable, $H(t)$ is the unit pulse component, β is the damping coefficient, and f_{RE} is the pulse resonance frequency.

Random noise $n(t)$ is an AWGN with SNR $= -8$db. The sampling frequency is set to 5120 Hz and the sampling time to 2 s. The main parameter values of hybrid simulation signal $y(t)$ are shown in Table 2.

Table 2 Main parameter value of hybrid simulation signal

$x(t)$				$r(t)$		$h(t)$	
Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
K	3	A_1, B_1	0.5	i	2	j	3
N	3	A_2, B_2	0.3	C_1	0.8	R_1	1
f_m	1000 Hz	A_3, B_3	0.1	C_2	0.5	R_2	1.5
f_s	50 Hz	α_{kn}	0	f_1	3 Hz	R_3	2
f_{sr}	30 Hz	β_{kn}	0	f_2	20 Hz	β	125
φ_k	0			θ_i	0	f_{RE}	2000 Hz

Fig. 3 Hybrid simulation signal
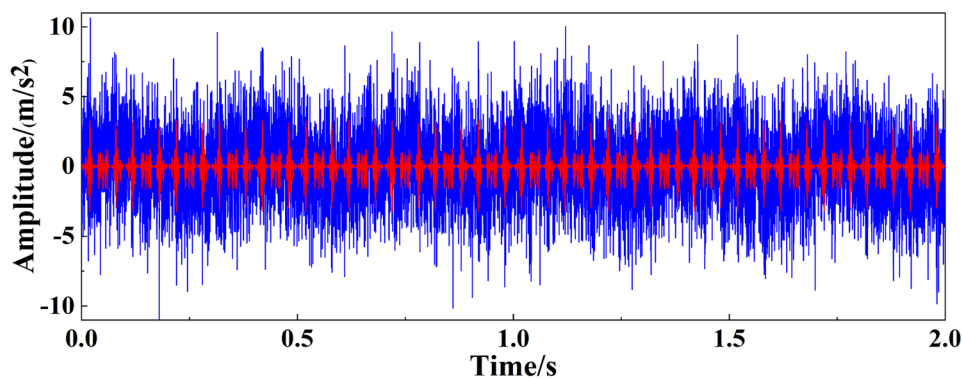
The time domain waveform of the hybrid simulation signal $y(t)$ is shown in Fig. 3, and the FFT spectrum and the envelope spectrum is shown in Fig. 4. Figure 3 shows that fault impact signal $x(t)$ has been completely covered by strong background noise, and the periodic impact caused by the fault cannot be observed in the time-domain waveform. In the FFT spectrum of hybrid simulation signal $y(t)$, only the frequencies (f_1 and f_2) of the two harmonic components can be observed, and fault frequency f_s cannot be observed. In the envelope spectrum of hybrid simulation signal $y(t)$, although the fault frequency f_s can be observed, its amplitude is not prominent, and there are many interference components around. During calculation, the search range of penalty factor α is set as [1000, 8000], the search range of decomposition layers K is set to [3, 7], the number of searched individuals is 15, and the maximum iteration number is 25. The optimal parameter combination calculated by the STOA is [5, 4100], and the iteration process is shown in Fig. 5. Convergence can be achieved after 10 iterations.

The decomposition result of the optimised VMD is shown in Fig. 6. In Fig. 6a, the IMF corresponding to the maximum value of *KEI* is IMF3, so IMF3 is selected as the sensitive component and envelope spectrum analysis is carried out. The analysis results are shown in Fig. 6b. The fault frequency and its frequency multiplier ($f_s = 50$ Hz, $2 fs = 1000$ Hz, $3 fs = 150$ Hz, $4 fs = 200$ Hz), the absolute rotation frequency of the sun gear ($f_{sr} = 30$ Hz) and related frequency (f_s-$f_{sr} = 20$ Hz) can be clearly extracted. The above results verify the effectiveness of the optimised VMD.

To further verify the superiority of the optimised VMD, this method is compared with EMD and EEMD. Among them, 9 IMFs were obtained by EMD, and 13 IMFs are obtained by EEMD. Figure 7 shows the envelope spectra of the first 4 IMFs by EMD and EEMD.

In Fig. 7a, EMD can extract absolute rotation frequency f_{sr}, fault frequency f_s and its 2 times frequency, but many interference components exist within the range of 100 Hz to 200 Hz. In Fig. 7b, EEMD can extract absolute rotation
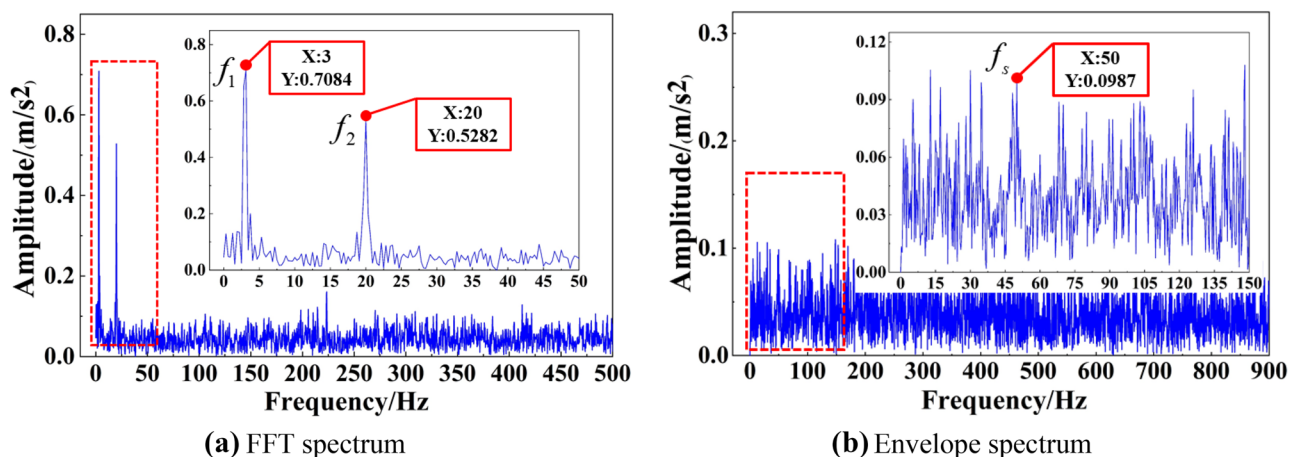

(a) FFT spectrum

(b) Envelope spectrum

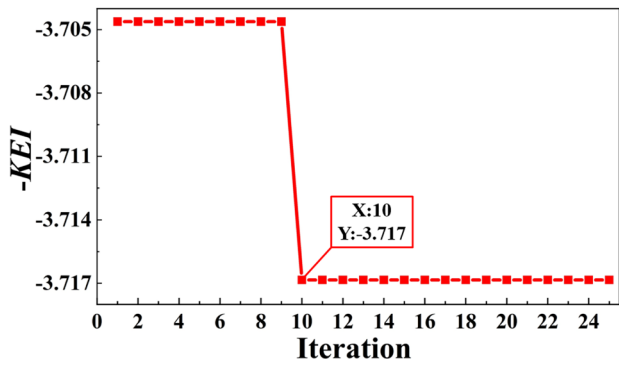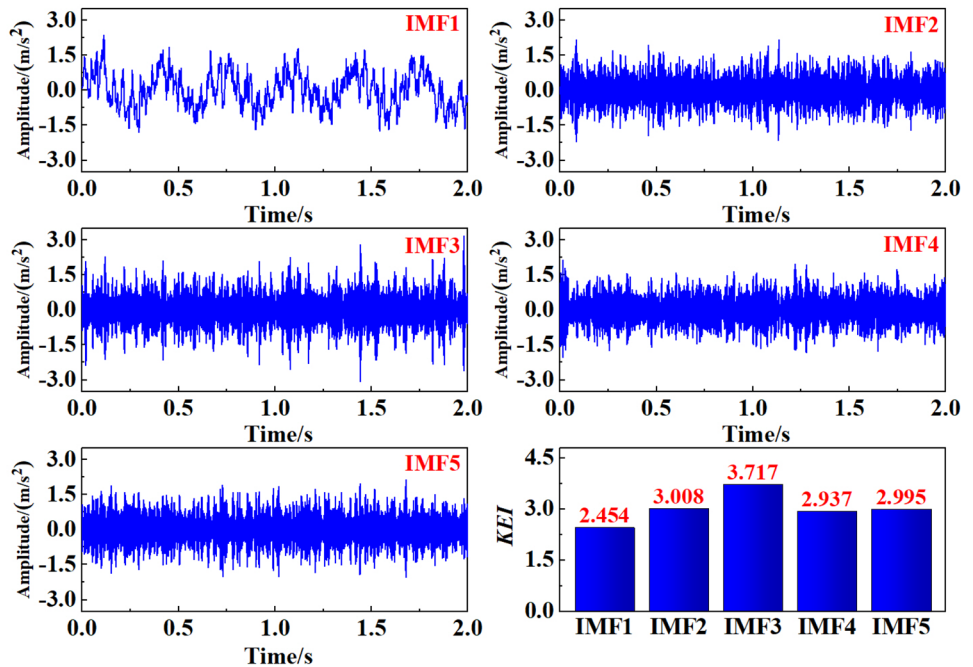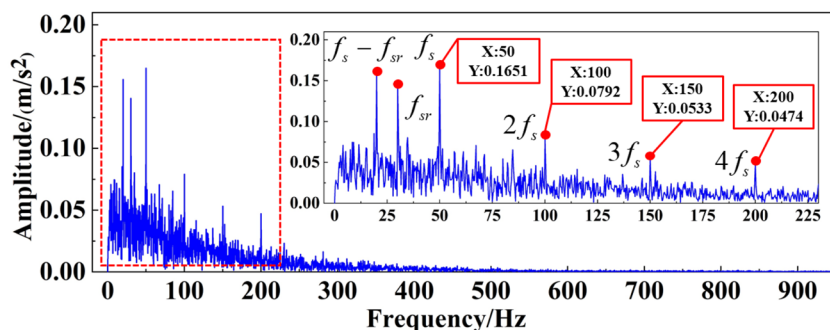**Fig. 4** FFT spectrum and envelope spectrum of hybrid simulation signal

Fig. 5 Iterative process

frequency f_{sr} and fault frequency f_s, but many interference components also exist within the range of 50 Hz to 200 Hz. The above results further verify the superiority of the optimised VMD.

5 Case analysis

5.1 Rotor unbalance fault experiment of motorized spindles

In this paper, the rotor unbalance fault simulation experiment of a certain type of motorized spindle was carried out by using the motorized spindle performance monitoring test platform. The structure of the test platform is shown in Fig. 8. The test platform is controlled by the control cabinet and the computer, which can simulate the actual working conditions of the motorized spindle. The tested motorized spindle is fixed by the clamping device, and the auxiliary devices such as oil–gas lubrication device, water cooler and frequency converter can ensure the operation of the tested electric spindle.

Rotor unbalance fault is simulated by adding different numbers of bolts at the output end of the tested

Fig. 6 Decomposition results of optimised VMD

(a) IMF1-IMF5

(b) Envelope spectrum of IMF3

Fig. 7 Decomposition results of EMD and EEMD

(a) EMD   **(b)** EEMD
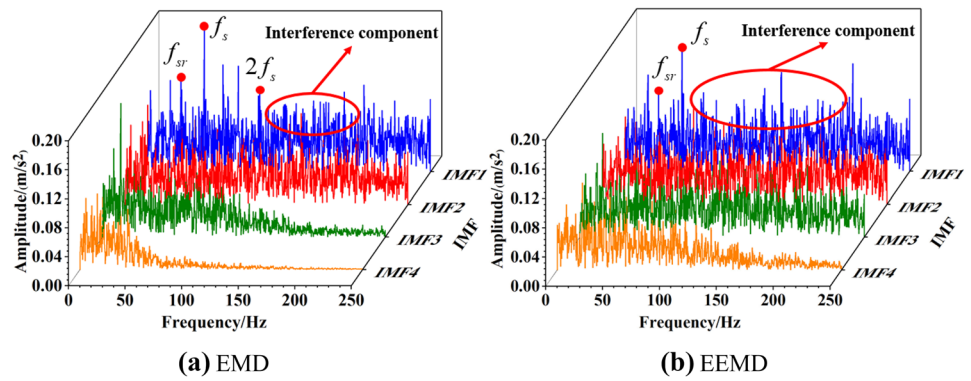
motorized spindle without applying force and torque. The health condition is divided into five levels. During the experiment, the vibration signal of the front bearing of the tested motorized spindle was monitored. The model of the acquisition device was NI-cDAQ9184, the spindle speed was 5000r/min, the sampling frequency was 1.28 kHz, the sampling time of the sample signal was 0.1 s, and 50 groups of sample signals were collected at each level.

According to Steps 2 to 4 in Sect. 3, the sample signal is decomposed by the optimised VMD, the *KEI* of the IMFs is calculated, and the selected IMFs is reconstructed. According to Step 5, the multi domain features of the reconstructed signal are extracted, and the normalised high-dimensional feature matrix is constructed. The mean value feature curves of all levels are shown in Fig. 9.

In Fig. 9, time domain features P_1, P_2, P_3, P_5 and frequency feature P_{11} can effectively identify each health condition. Although the entropy features (P_{17} to P_{22}) in the scale domain can accurately identify levels 4 and 5,

the recognition effect for levels 1 to 3 is poor. To sum up, if all 25-dimensional features are used to represent the health condition of the motorized spindle, redundant information will be inevitably produced. According to Step 6, SL-Isomap was used to carry out secondary feature extraction to construct low-dimensional feature matrix *F′*. In SL-Isomap, the nearest neighbour parameter is set to 50 and regulating factor *a* to 0.5. The visualisation result of a low-dimensional feature matrix is shown in Fig. 10. In Fig. 10, intrinsic dimension $d = 3$, the samples of each health condition can be basically stripped, and the spacing between the samples is clear.

The training and test samples were randomly divided in 2:3. Ready, the number of hidden states of the GMM-HMM was set to four, the number of Gaussian components to two, and the maximum number of iterations to 50. The iteration process is shown in Fig. 11. The curves of all the levels converge after the 24th, 11th, 16th, 22nd, and 8th iterations. The library is constructed based on GMM-HMM, and then the test samples are respectively input into the library to calculate LLP. The health condition

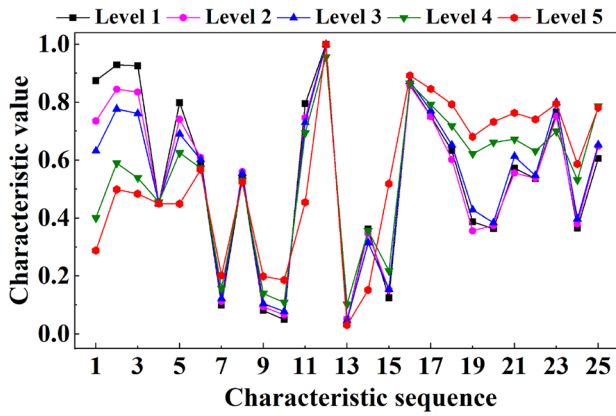**Fig. 8** Motorized spindle performance monitoring test platform
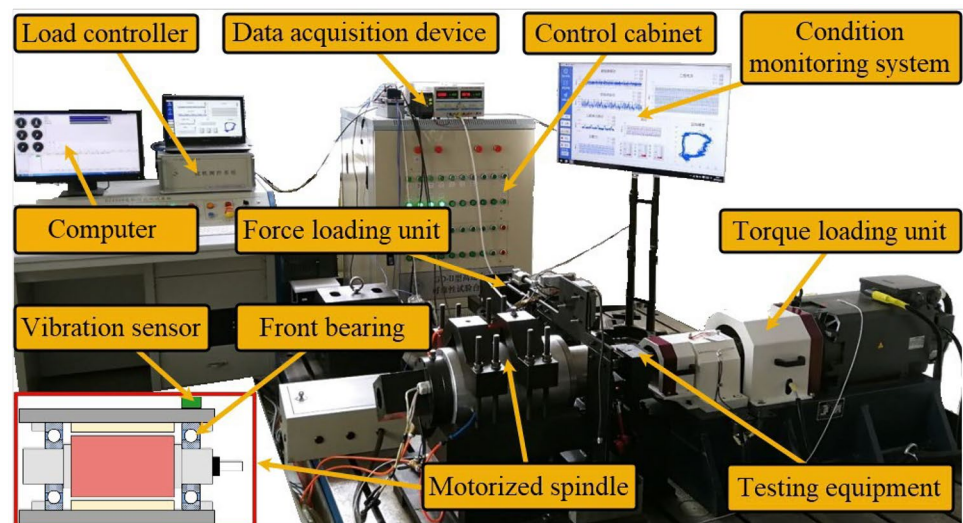
Fig. 9 Mean feature curves of all levels

Fig. 11 Training curves of GMM-HMM

corresponding to the maximum LLP is the most likely condition of the motorized spindle.

The evaluation results of five health condition levels are shown in Fig. 12 in which solid points represent the real labels of the test samples. The label corresponding to the highest point in each sample is the evaluation result. In Fig. 12, the evaluation accuracy of level 1 is 90%, level 2 is 90%, level 3 is 86.67%, level 4 is 90%, level 5 is 86.67%, and the average evaluation accuracy of the five levels is 88.67%. When the evaluation result is wrong, the evaluation label and the real label are always adjacent, which also conforms to the degradation law of motorized spindle. The above results verify the effectiveness of the proposed method.
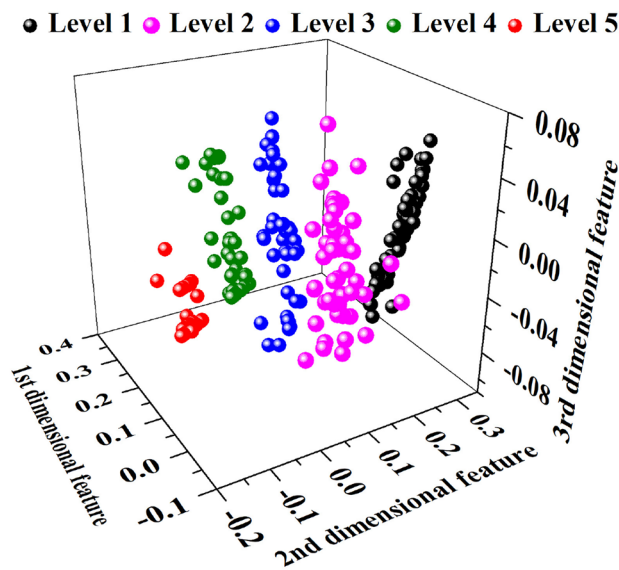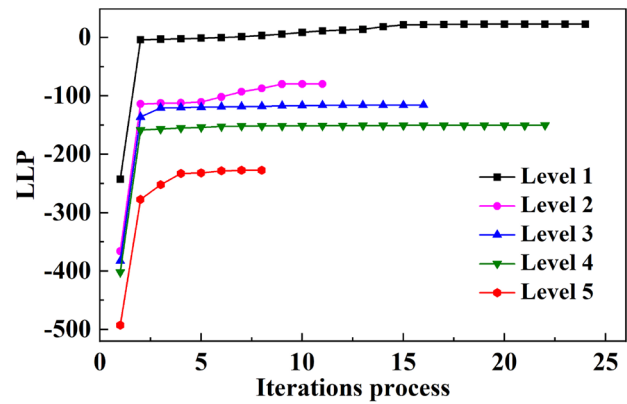
5.2 Comparison and analysis

The 'SL-ISOMAP + GMM-HMM' was compared with several combination methods to further verify the superiority of the method. Various combination methods are composed of three manifold learning methods and two classical classifiers [31, 32]. The three manifold learning methods include Isomap, t-Distributed Stochastic Neighbour Embedding (t-SNE), and Laplacian eigenmaps (LE). The two classical classifiers include PSO–SVM and genetic algorithm optimisation-ELM (GA-ELM).

The nearest neighbour parameter of the Isomap, the LLE and the LE is set to 50, and the perplexity of the t-SNE to 30. In PSO–SVM, the population size is 20, the learning factor is 1.5, the inertia weight is 0.6, the maximum number of iterations is 50, and the radial basis function is selected as the kernel function. In GA–ELM, the population size is 20, the generation gap is 0.95, the crossover probability is 0.7, the mutation probability is 0.01, the maximum genetic generation number is 50 and the number of neurons in the hidden layer is 30. Taking the average accuracy as the performance index, the calculation results of various combination methods are shown in Table 3.

In Table 3, the average accuracy of 'SL-Isomap + GMM-HMM' ranks first among the combination methods. Compared with 'SL-Isomap + PSO-SVM' and 'SL-Isomap + GA-ELM', the average accuracy of the proposed method improved by 4% and 5.33%, respectively. For the same manifold learning method, the average accuracy of GMM-HMM is the highest, and that of GA-ELM is the worst. For the same classifier, SL-Isomap has the best dimension reduction performance, while LE has the worst. For levels 3, 4 and 5, the accuracy

Fig. 10 Visualisation of low-dimensional feature matrix

Fig. 12 Evaluation results of five health condition levels

(a) Health condition level 1

(b) Health condition level 2

(c) Health condition level 3

(d) Health condition level 4

(e) Health condition level 5
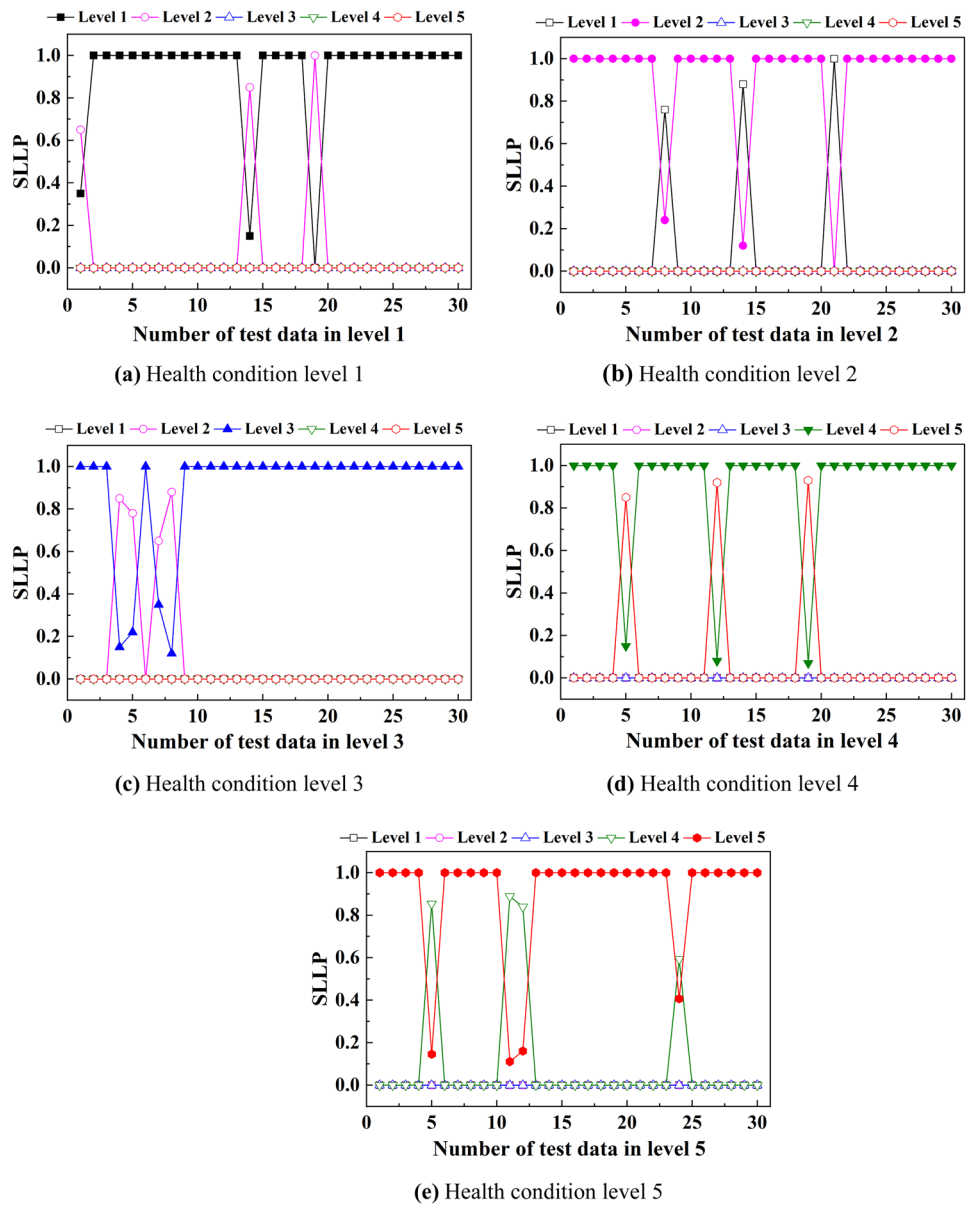
Table 3 Evaluation accuracy of combined methods

Classification algorithm	Manifold learning	Accuracy(%)					
		Level 1	Level 2	Level 3	Level 4	Level 5	Average
GMM-HMM	SL-Isomap	90.00	90.00	86.67	90.00	86.67	88.67
	Isomap	73.33	63.33	66.67	73.33	80.00	71.33
	t-SNE	83.33	80.00	76.67	70.00	66.67	75.33
	LE	50.00	56.67	53.33	56.67	46.67	52.67
PSO-SVM	SL-Isomap	80.00	90.00	83.33	86.67	83.33	84.67
	Isomap	63.33	66.67	63.33	56.67	60.00	62.00
	t-SNE	80.00	73.33	66.67	76.67	66.67	72.67
	LE	53.33	56.67	46.67	56.67	36.67	50.00
GA-ELM	SL-Isomap	90.00	83.33	80.00	83.33	80.00	83.33
	Isomap	66.67	53.33	63.33	50.00	56.67	58.00
	t-SNE	63.33	66.67	73.33	63.33	70.00	67.33
	LE	53.33	46.67	56.67	46.67	43.33	49.33

of 'SL-Isomap + GMM-HMM' is higher than the other combined methods. For level 1, the accuracy of 'SL-Isomap + GMM-HMM' and 'SL-Isomap + GA-ELM' is 90%. For level 2, the accuracy of 'SL-Isomap + GMM-HMM' and 'SL-Isomap + PSO-SVM' is 90%. The above results further verify the superiority of 'SL-Isomap + GMM-HMM'.

6 Conclusion

Extracting and using the useful information of vibration signals in the industrial environment with a strong noise are the key problem of the health condition evaluation of motorized spindle. In this paper, a health condition evaluation method for motorized spindles on the basis of the optimised VMD and the GMM-HMM is proposed. This paper is summarised as follows:

1. The optimised VMD is used to decompose the hybrid simulation signal, and the fault frequency and its frequency doubling part can be clearly extracted from the envelope spectrum of the sensitive IMF. The results show that the optimised VMD can effectively extract the fault information hidden in strong background noise.
2. To verify the proposed method, rotor unbalance experiment was carried out on a certain type of motorized spindle. The evaluation results show that the accuracy of levels 1 to 5 is 90%, 90%, 86.67%, 90%, and 86.67%, respectively, and the average accuracy is 88.67%, indicating that the proposed method can effectively evaluate the health condition of motorized spindle.
3. The average accuracy of 'SL-Isomap + GMM-HMM' ranks first among the combination methods. The result further verifies the superiority of the proposed method.

Author contribution Haiji Yang: Background research, Data curation, Software, Validation writing-original draft, Editing. Guofa Li: Methodology, Review & editing, Supervision. Jialong He: Project administration, Funding acquisition. Liding Wang: Supervision. Xinyu Nie: Assist in experiment. All authors read and approved the final manuscript.

Data availability Not applicable.

Code availability Not applicable.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

References

1. Liu ZF, Pan MH, Zhang AP (2015) Thermal characteristic analysis of high-speed motorized spindle system based on thermal contact resistance and thermal-conduction resistance. Int J Adv Manuf Tech 76(9–12):1913–1926. https://doi.org/10.1007/s00170-014-6350-1
2. Yang Z, Li X, Chen C (2019) Reliability assessment of the spindle systems with a competing risk model. Proc Inst Mech Eng O-J Risk Reliab 233(2):226–234. https://doi.org/10.1177/1748006X18770343
3. Zhang Z, Cheng Q, Qi B (2021) A general approach for the machining quality evaluation of S-shaped specimen based on POS-SQP algorithm and Monte Carlo method. Manuf Syst 60:553–568. https://doi.org/10.1016/j.jmsy.2021.07.020
4. Cheng Q, Qi B, Liu Z (2019) An accuracy degradation analysis of ball screw mechanism considering time-varying motion and loading working conditions. Mech Mach Theory 134:1–23. https://doi.org/10.1016/j.mechmachtheory.2018.12.024
5. Moumene I, Ouelaa N (2022) Gears and bearings combined faults detection using optimized wavelet packet transform and pattern recognition neural networks. Int J Adv Manuf Tech 120(7–8):4335–4354. https://doi.org/10.1007/s00170-022-08792-2
6. Song Y, Zeng S, Ma J (2018) A fault diagnosis method for roller bearing based on empirical wavelet transform decomposition with adaptive empirical mode segmentation. Measurement 117:266–276. https://doi.org/10.1016/j.measurement.2017.12.029
7. Liu J, Gu Y, Chou Y (2021) Seismic data random noise reduction using a method based on improved complementary ensemble EMD and adaptive interval threshold. Explor Geophys 52(2):137–149. https://doi.org/10.1080/08123985.2020.1777849
8. Gupta P, Singh B (2022) Ensembled local mean decomposition and genetic algorithm approach to investigate tool chatter features at higher metal removal rate. J Vib Control 28(1–2):30–44. https://doi.org/10.1177/1077546320971157
9. Fei S-w (2017) Fault diagnosis of bearing based on wavelet packet transform-phase space reconstruction-singular value decomposition and SVM classifier. Arab J Sci Eng 42(5):1967–1975. https://doi.org/10.1007/s13369-016-2406-x
10. Sun Y, Chen H, Shi Z (2020) A novel bevel gear fault diagnosis method based on ensemble empirical mode decomposition and support vector machines. Insight 62(1):34–41. https://doi.org/10.1784/insi.2020.62.1.34
11. Tian Y, Ma J, Lu C (2015) Rolling bearing fault diagnosis under variable conditions using LMD-SVD and extreme learning machine. Mech Mach Theory 90:175–186. https://doi.org/10.1016/j.mechmachtheory.2015.03.014
12. Zhu Y, Yan Q, Lu J (2020) Fault diagnosis method for disc slitting machine based on wavelet packet transform and support vector machine. Int J Comput Integr Manuf 33(10–11):1118–28. https://doi.org/10.1080/0951192X.2020.1795927
13. Wang J, Li J, Wang H (2021) Composite fault diagnosis of gearbox based on empirical mode decomposition and improved variational mode decomposition. J Low Freq Noise Vib Act Control 40(1):332–346. https://doi.org/10.1177/1461348420908364
14. Zhang Y, Ji J, Ma B (2020) Fault diagnosis of reciprocating compressor using a novel ensemble empirical mode

decomposition-convolutional deep belief network. Measurement. https://doi.org/10.1016/j.measurement.2020.107619

15. Dragomiretskiy K, Zosso D (2014) Variational mode decomposition. IEEE Trans Signal Process 62(3):531–544. https://doi.org/10.1109/TSP.2013.2288675

16. Li F, Li R, Tian L (2019) Data-driven time-frequency analysis method based on variational mode decomposition and its application to gear fault diagnosis in variable working conditions. Mech Syst Signal Process 116:462–479. https://doi.org/10.1016/j.ymssp.2018.06.055

17. An X, Zhang F (2017) Pedestal looseness fault diagnosis in a rotating machine based on variational mode decomposition. Proc Inst Mech Eng C-J Mech 231(13):2493–2502. https://doi.org/10.1177/0954406216637378

18. Kumar A, Zhou Y, Xiang J (2021) Optimization of VMD using kernel-based mutual information for the extraction of weak features to detect bearing defects. Measurement. https://doi.org/10.1016/j.measurement.2020.108402

19. Zhong J, Gou X, Shu Q, Liu X, Zeng Q (2021) A FOD detection approach on millimeter-wave radar sensors based on optimal VMD and SVDD. Sensors 21(3). https://doi.org/10.3390/s21030997

20. Dhiman G, Kaur A (2019) STOA: A bio-inspired based optimization algorithm for industrial engineering problems. Eng Appl Artif Intell 82:148–174. https://doi.org/10.1016/j.engappai.2019.03.021

21. Mor B, Garhwal S, Kumar A (2021) A systematic review of hidden markov models and their applications. Arch Comput Method Eng 28(3):1429–1448. https://doi.org/10.1007/s11831-020-09422-4

22. Dong L, Li W-m, Wang C-H (2020) Gyro motor fault classification model based on a coupled hidden Markov model with a minimum intra-class distance algorithm. Proc Inst Mech Eng I-J Syst Control Eng 234(5):646–661. https://doi.org/10.1177/0959651819866281

23. Cheng G, Li H, Hu X (2017) Fault diagnosis of gearbox based on local mean decomposition and discrete hidden Markov models. Proc Inst Mech Eng C-J Mech 231(14):2706–2717. https://doi.org/10.1177/0954406216638885

24. Liu T, Chen J, Dong G (2016) Identification of bearing faults using linear discriminate analysis and continuous hidden Markov model. Proc Inst Mech Eng C-J Mech 230(10):1658–1672. https://doi.org/10.1177/0954406215582015

25. Zheng B, Gao X (2017) Sucker rod pumping diagnosis using valve working position and parameter optimal continuous hidden Markov model. J Process Control 59:1–12. https://doi.org/10.1016/j.jprocont.2017.09.007

26. Xia J, Shang P, Wang J (2016) Permutation and weighted-permutation entropy analysis for the complexity of nonlinear time series. Commun Nonlinear Sci 31(1–3):60–68. https://doi.org/10.1016/j.cnsns.2015.07.011

27. Yin Y, Shang PJ (2017) Multivariate weighted multiscale permutation entropy for complex time series. Nonlinear Dynam 88(3):1707–1722. https://doi.org/10.1007/s11071-017-3340-5

28. Tenenbaum JB, de Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290(5500):2319-+. https://doi.org/10.1126/science.290.5500.2319

29. Geng X, Zhan DC, Zhou ZH (2005) Supervised nonlinear dimensionality reduction for visualization and classification. IEEE Trnas Syst Man Cybern B Cybern 35(6):1098–1107. https://doi.org/10.1109/TSMCB.2005.850151

30. Zhang M, Wan KS, Wei DD (2018) Amplitudes of characteristic frequencies for fault diagnosis of planetary gearbox. J Sound Vib 432:119–132. https://doi.org/10.1016/j.jsv.2018.06.011

31. Zheng JD, Jiang ZW, Pan HY (2018) Sigmoid-based refined composite multiscale fuzzy entropy and t-SNE based fault diagnosis approach for rolling bearing. Measurement 129:332–342. https://doi.org/10.1016/j.measurement.2018.07.045

32. An J, Ai P, Liu C (2021) Deep clustering bearing fault diagnosis method based on local manifold learning of an autoencoded embedding. IEEE Access 9:30154–30168. https://doi.org/10.1109/ACCESS.2021.3059459