Research Article

Research on the Fault Diagnosis Method for Rolling Bearings Based on Improved VMD and Automatic IMF Acquisition

Ying Zhang [1,2] **and Anchen Wang**[1]

¹College of Automobile and Traffic Engineering, Nanjing Forestry University, Nanjing 210037, China
*²Key Laboratory of Civil Aviation Aircraft Health Monitoring and Intelligent Maintenance,
 Nanjing University of Aeronautics and Astronautics, Nanjing 210037, China*

Correspondence should be addressed to Ying Zhang; zhangyingrms@163.com

This paper proposes a novel method to improve the variational mode decomposition (VMD) method and to automatically acquire the sensitive intrinsic mode function (IMF). First, since fault signals are impulsive and periodic, a weighted autocorrelative function maximum (AFM) indicator is constructed based on the Gini index and autocorrelation function to serve as the optimization objective function. The mode number K and the penalty parameter α of VMD are automatically obtained through an optimal parameter searching process underpinned by the improved particle swarm optimization (PSO) algorithm with a variety of inertia weights. This improvement solves one of the major drawbacks of the conventional VMD method, that is, the need to manually set parameters. Then, an optimal IMF automatic selecting process is performed for single-failure faults and compound faults, according to the principles of the maximum weighted AFM indicator and maximum spectrum peak ratio (SPR), respectively. The sensitive IMFs are then subjected to an envelope demodulation analysis to obtain the fault characteristic frequency. The results of simulations and experiments show that the proposed method can effectively identify fault characteristics early, especially compound faults, demonstrating great potential for real-world applications.

1. Introduction

Developing techniques for monitoring rolling bearing conditions and diagnosing faults are key to realizing the transition from traditional regular maintenance to condition-based maintenance. Owing to harsh working environments and changing conditions, vibrations of rolling bearings are typically characterized by strong background noise, unsteadiness, and coupling modulation, which make it difficult to extract and identify fault characteristics. Nonetheless, accurately extracting fault characteristic frequencies from a mix of complex, nonstationary, and nonlinear vibration signals is essential for early fault identification [1].

Time-frequency analysis methods simultaneously provide local information about nonstationary signals in both the time domain and frequency domain and can be used to extract fault characteristics from weak signals. Some commonly used time-frequency analysis methods include empirical mode decomposition (EMD) [2], ensemble empirical mode decomposition (EEMD) [3], and local mean decomposition (LMD) [4]. All of these are based on recursive decomposition and, thus, have the same disadvantages, such as endpoint effects and modal aliasing, which can impact accuracy in early detection of weak fault characteristics [5].

Dragomirestkiy and Zosso [6] recently proposed the VMD method, which is based on solid mathematical foundations and can be used for broad applications in the field of mechanical fault diagnosis. The VMD method introduces a secondary penalty factor and a Lagrange multiplier to transform the constrained variational problem into a nonconstrained variational problem. This nonrecursive decomposition approach avoids some of the problems of recursive decomposition, such as the endpoint effect and the

need to determine conditions for terminating the decomposition, and, in addition, effectively remedies the deficiencies of conventional decomposition methods and achieves high computational efficiency and good noise robustness. However, the VMD method also has a number of drawbacks. Most notably, the mode number K and the penalty parameter α must be set before starting the decomposition. Moreover, problems like aliasing between mode components or excessive decomposition may occur if parameters are not set properly. Therefore, adaptive acquisition of VMD parameters is of crucial importance.

To tackle this challenge, Shi and Yang [7] proposed a step-by-step parameter optimization strategy which achieved good results in validation experiments using a wind turbine. Wang et al. [8] used the minimum average envelope spectrum entropy as the objective function and PSO to optimize the VMD parameters. The method was shown to be superior to EMD, LMD, and other envelope and wavelet transform methods. Yan et al. [9] proposed a composite fault detection method for rotating machinery based on optimal variational mode decomposition (OVMD) and a 1.5-dimensional envelope spectrum. The envelope spectrum entropy value was used as the objective function, and the optimal VMD parameters were obtained using the genetic algorithm, thereby achieving adaptive acquisition of the mode number and penalty parameter. In another approach, Zhang et al. [10] proposed a grasshopper optimization algorithm- (GOA-) based adaptive VMD parameter analysis method that works in the following way: a weighted kurtosis indicator is constructed based on the kurtosis and correlation coefficient; VMD parameters are optimized using the GOA algorithm with the weighted kurtosis indicator as the objective function; and the sensitive IMF is selected following the principle of maximum weighted kurtosis for extracting the fault characteristics. Liu et al. [11] used the principle of maximum kurtosis to obtain the optimal values of parameter pair K and α in order to achieve an accurate VMD. Lian et al. [12] proposed an adaptive VMD method that uses the principle of maximum kurtosis as an evaluation criterion to automatically determine the mode number.

Based on the above studies, it can be concluded that constructing a suitable objective function and selecting an appropriate optimization algorithm are two critical issues in the adaptive acquisition of VMD parameters. Some indicators such as kurtosis [10–14], correlation coefficient [15], autoregressive (AR) model parameters [16], and entropy [8] have proven extremely useful in vibration signal analysis and fault characteristics extraction. As a sparse indicator, kurtosis is widely used in the fault diagnosis of rotating machinery and can effectively extract impulsive components of signals. The principle of maximum kurtosis is useful for constructing the objective function for VMD parameter optimization. However, kurtosis is more suitable for measuring the most important pulses in a signal. In addition to not being robust enough, kurtosis-based techniques focus more on individual pulses than on the desired periodic pulses [17].

The Gini index is widely used to measure income inequality in countries and has recently been used in fault diagnosis of rotating machinery [18]. The index is not only very effective in identifying fault pulses but offers superb performance in preventing random pulses and disturbances. Miao et al. [19] used the Gini index to improve the kurtosis guidance map and were able to identify the resonance band, which contains the richest information. More specifically, the resonance band contains the fault pulse rather than interferences caused by fault modulation and random pulses. Kurtosis and the Gini index are both used as indicators but share the same shortcoming; that is, both measures are aimed at detecting the impact of fault pulses and ignore recurring periodic pulses.

The autocorrelation function retains periodic signal information; however, phase information of the original signal is not retained. For white noise signals, the autocorrelation function quickly decays to zero at the fastest decay rate. For an impulse signal, the autocorrelation function also acquires impulse characteristics while retaining periodic characteristics. The autocorrelation function of a noise-containing impulse signal decays rapidly, but its amplitude becomes large during the impulse. This suggests that the autocorrelation function of an impulse signal combines characteristics of the autocorrelation function of white noise and impulse signal. Chen et al. [20] proposed an AFM indicator based on this particular feature of the autocorrelation function, which was then used as a quantitative evaluation criterion of the IMF value after complementary EEMD processing and achieved good results.

Up to now, most studies in the literature have focused on single-failure faults (e.g., inner ring failure, outer ring faults, and rolling element faults of rolling bearings). In practice, most faults are compound and composed of two or more different types of faults. At present, there are many research studies on the diagnosis of compound faults. Wang et al. [21] proposed a modified VMD (MVMD) based on noise-assisted data analysis (NADA), which can extract compound fault features of gearbox under a strong noise environment. Yang et al. [22] proposed a fault diagnosis framework based on a novel dual-extreme learning machine (ELM) network that can quickly detect single and simultaneous failures in a multistage gearbox. Zhong et al. [23] proposed a fault feature extraction framework based on feature extraction and support vector machine (SVM), and the experimental results show that the method can effectively diagnose single and simultaneous faults of the gas turbine generator system (GTGS). Wang et al. [24] proposed a gearbox integrated fault diagnosis method based on ensemble LMD, defined as CELMD (Complementary Ensemble Local Mean Decomposition) which overcomes the pattern mixing phenomenon and can be used to extract composite fault features. Wang [25] proposed a method based on Multipoint Optimal Minimum Entropy Deconvolution Adjusted (MOMEDA) and EEMD, which is suitable for searching complex fault pulse signals under strong noise environment. Wang et al. [26] proposed a maximum kurtosis spectrum deconvolution (MKSED) applied to the fault diagnosis of the gearbox, and the experimental results show that the compound faults of gearbox can be successfully diagnosed by using MKSED.

In this paper, a method for improving the VMD is proposed; a new weighted AFM indicator is constructed to serve as the objective function for parameter optimization. The improved PSO with a variety of inertia weights is used for optimal parameter searching. Mutual information of the decomposed components and the original signal is used to detect overdecomposition, upon which the decision to perform the search again is made. The VMD is performed after the optimal decomposition parameters are obtained. For single-failure faults, the weighted AFM indicator is used to select the sensitive IMF and for compound faults, SPR is used. Finally, the selected IMF undergoes an envelope demodulation analysis to extract the fault characteristic frequency. The following sections are arranged as follows. The VMD method is introduced in Section 2. The construction of weighted WAFM index, IPSO algorithm, mutual information, and the method proposed in this paper are introduced in Section 3. The simulation signal and test signal are analyzed using the proposed method, respectively, in Section 4 and Section 5. Finally, the conclusions of this paper are drawn in Section 6.

2. VMD

The VMD is an adaptive signal processing technique first proposed in 2014 that can be used to decompose a signal into a series of finite-bandwidth IMFs in a nonrecursive manner, thereby creating the right conditions for effective signal separation [6]. Migrating the signal decomposition process into the variational framework, VMD accomplishes signal decomposition by seeking the optimal solution of the variational model.

2.1. Construction of the Variational Problem.
With VMD, the signal is decomposed into K mode functions, each with a center frequency of ω_k. Therefore, the VMD can be expressed as the variational problem of solving K mode functions whose aggregated bandwidth is the smallest, constrained by the sum of all mode functions which must be equal to the original signal:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}$$

$$\text{s.t.} \qquad \sum_l u_k = f, \tag{1}$$

where $\{\omega_k\} = \{\omega_1, \omega_2, \ldots, \omega_k\}$ represents the center frequency of each mode.

The VMD can be obtained using the following procedure.

The Hilbert transform is performed on each mode function u_k to obtain the analytical signal:

$$\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t). \tag{2}$$

Then, a preliminary estimation is performed on the center frequency $e^{-j\omega_k t}$ of the analytical mixed signal of each mode. The frequency of each mode component is modulated by its corresponding baseband frequency:

$$\left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t}. \tag{3}$$

The square of the gradient 2-norm of the above demodulated signal is calculated and the bandwidth of each modal function is estimated.

2.2. Solution to the Variational Problem.
To solve the variational problem presented in Section 2.1, the secondary penalty factor α and Lagrange multiplier $\lambda(t)$ are introduced and used to transform the constrained variational problem into an unconstrained variational problem. The secondary penalty factor serves as a guarantee of the accuracy of signal reconstruction when Gaussian noise is present, while the Lagrange multiplier makes the constraint more stringent. The augmented Lagrange expression is as follows:

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

$$+ \left\| f(t) - \sum_k u_k(t) \right\| \tag{4}$$

$$+ \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle.$$

Both the IMF and its center frequency are continuously updated using the alternate direction method of multipliers (ADMM). Finally, the saddle point of the above formula is obtained, and this is the optimal solution. The IMF of all frequency domains along with their center frequencies can be obtained using the following formulae:

$$\hat{u}_k^{n+1}(w) = \hat{f}(w) - \frac{\sum_{i \neq k} \hat{u}_i(w) + (\hat{\lambda}(w)t/2)}{1 + 2\alpha(\omega - \omega_k)}, \tag{5}$$

$$\hat{\omega}_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}(w)|^2 d\omega}{\int_0^\infty |\hat{u}(w)|^2 d\omega}, \tag{6}$$

where $\hat{u}_k^{n+1}(w)$ is the result of the current residual amount $\hat{f}(w) - \sum \hat{u}_i(w)$ after Wiener filtering.

The above is an adaptive decomposition procedure of VMD, which is in essence a Wiener filtering process. Signal decomposition is realized by iteratively searching for the optimal solution of the variational model in the variational problem framework. The approach achieves good noise robustness and effectively avoids the endpoint effect and modal aliasing problem.

3. WAFM-IPSO-VMD

To eliminate the need to set parameters in advance, which is a drawback of the VMD, this paper proposes a method for improving the VMD by enabling adaptive parameter acquisition. The method includes three aspects: first, the weighted AFM indicator is constructed as the objective function; second, the improved PSO is used to search for the

optimal parameters of the VMD algorithm and mutual information of the decomposed components, and the original signal is used to detect overdecomposition, and thus to decide whether or not to perform the search again; third, an optimal IMF automatic selecting process is performed for the single-failure fault and compound faults following the principles of the maximum weighted AFM and maximum SPR, respectively.

3.1. WAFM.

When searching for the optimal parameter combination using the improved PSO, construction of the objective function is a critical step. If a fault occurs, a periodic impulse signal will appear in the mixed signal. As described before, the Gini index aims to detect the impact of fault pulses and ignores periodic pulses that repeatedly occur when a fault is present, whereas AFM can effectively extract periodic characteristics of the signal. To enable more effective fault detection, this paper proposes constructing the weighted AFM indicator based on characteristics of the Gini index and the AFM.

3.1.1. Gini Index.

The Gini index is an inequality index that is often used in economics to reflect the overall income gap [27]. In mechanical fault diagnosis, the sequence elements must be positive in order to calculate the Gini index; therefore, the envelope spectrum Gini index has been used to extract the bearing fault signal [19] and achieved good results. The envelope spectrum Gini index can be defined as follows:

$$\text{Gini} = 1 - 2 \sum_{n=1}^{N} \frac{x_{(n)}}{\|\vec{x}\|} \left(\frac{N - n + (1/2)}{N} \right), \tag{7}$$

where the vector $\vec{x}$ is the envelope spectrum of the signal; $\vec{x} = [x_{(1)}, x_{(2)}, \ldots, x_{(N)}]$ is a sequence of points in the envelope spectrum arranged in ascending order of amplitude; and $(1), (2), (3), \ldots, (N)$ are the new index numbers after sorting.

$0 \leq \text{Gini} \leq 1$. In economics, the closer the Gini index is to zero, the more equal the income distribution is; the closer the Gini index is to 1, the more uneven the income distribution is. Thus, extracting a bearing fault signal from the vibration signal of a faulty machine by applying the principle of the maximum Gini index of the envelope spectrum is feasible.

3.1.2. AFM.

The autocorrelation function is extremely effective in ascertaining the periodicity of a signal and can determine whether the signal is noisy [20].

The autocorrelation function describes how similar a signal is to itself at different times. For a random process $\{x_t\}$, a signal that occurs at a certain moment in time is denoted x_t and the same signal becomes x_{t+k} after a certain period k. Self-covariance of the signal can be defined as

$$y_t = \text{cov}[x_t, x_{t+k}] = E[(x_t - \overline{x})(x_{t+k} - \overline{x})]. \tag{8}$$

The autocorrelation coefficient of x_t and x_{t+k} is

$$\rho_k = \frac{E[(x_t - \overline{x})(x_{t+k} - \overline{x})]}{\sqrt{\text{Var}(x_t)}\sqrt{\text{Var}(x_{t+k})}}. \tag{9}$$

For a stationary process, $\text{Var}(x_t) = \text{Var}(x_{t+k}) = \sigma_x^2$. So, the autocorrelation coefficient of x_t and x_{t+k} can be defined as

$$\rho_k = \frac{E[(x_t - \overline{x})(x_{t+k} - \overline{x})]}{\sigma_x^2}. \tag{10}$$

The autocorrelation coefficient sequence $\rho_k, k = 0, 1, \ldots, k$ with the time interval of k serving as a variable is called the autocorrelation function and can be defined as

$$R_{xx}(k) = \frac{1}{T\sigma_x^2} \sum_{t=1}^{T-k} (x_t - \overline{x})(x_{t+k} - \overline{x}). \tag{11}$$

The autocorrelation function is an even function. When performing the analysis, it is enough to analyze one-half of the autocorrelation function corresponding to the positive half of the axis. In signal processing, the autocorrelation function is defined as

$$R_x(k) = \frac{1}{T} \sum_{t=1}^{T-k} x_t \cdot x_{t+k}. \tag{12}$$

As a sequence, the autocorrelation function cannot describe the characteristics of the signal. To convert the autocorrelation function into a quantitative indicator, the root mean square value of the maximum value sequence (denoted as AFM [20]) of the autocorrelation function can be used as an index for evaluating the IMF. The AFM can be expressed by the following formula:

$$\text{AFM} = \sqrt{\sum_{n=1}^{N} \frac{P_n^2}{N}}, \tag{13}$$

where P_n is the maximum point sequence of the autocorrelation function of the signal envelope and N is the sequence length.

3.1.3. WAFM.

In view of the impulsive and periodic nature of fault signals, this paper uses the product of the Gini index and the AFM as the objective function for VMD parameter optimization. Since $0 \leq \text{Gini} \leq 1$, multiplying the AFM by the Gini index is equivalent to multiplying the AFM by a weight. Hence, the new indicator is defined as the weighted AFM:

$$\text{WAFM} = \text{Gini} \cdot \text{AFM}. \tag{14}$$

3.2. Improved PSO.

The PSO algorithm is a population-based random search method proposed by Kennedy and Eberhart [28]. The algorithm can be detailed as follows:

Suppose that the population space dimensions are $N \times D$, where N represents the number of populations, D represents the dimensionality of particles; the D-

dimensional vector $X_i = (x_{i1}, x_{i2}, \ldots, x_{iD})$ represents the position $(i = 1, 2, \ldots, N)$ of the ith particle and its velocity can be expressed as $V_i = (v_{i1}, v_{i2}, \ldots, v_{iD})$; $p_{\text{best}} = (p_{i1}, p_{i2}, \ldots, p_{iD})$ represents the individual extremum; and $g_{\text{best}} = (g_1, g_2, \ldots, g_D)$ represents the global extremum. After the individual and global extrema are obtained, the particle velocity and position are updated using the following formulae:

$$v_{ij}(t+1) = wv_{ij}(t) + c_1 r_1 \left[p_{ij}(t) - x_{ij}(t) \right]$$
$$+ c_2 r_2 \left[p_{gi}(t) - x_{ij}(t) \right], \tag{15}$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1), \tag{16}$$

where c_1 and c_2 are learning factors, r_1 and r_2 are random numbers uniformly distributed within $[0, 1]$, and w is the inertia weight.

The inertia weight affects local convergence and global exploration ability of the algorithm. Global exploration is performed when the weight is large and a local search is carried out when the weight is small. The PSO algorithm with a nonlinear decrease of the inertia weight improves upon on the inertia weight w [29]. The improved PSO algorithm is more effective since the solution space is analyzed during the initial search period such that the algorithm quickly approaches the optimal solution. In the late stage of search, higher accuracy can be achieved through iterative fine-tuning. The formula for improving the weight is as follows:

$$w = \left(\frac{\text{iter}_{\max} - \text{iter}}{\text{iter}_{\max}} \right)^n (w_{\max} - w_{\min}) + w_{\min}, \tag{17}$$

where $w_{\max}$ is the inertia weight after initialization, $w_{\min}$ is the final inertia weight, $\text{iter}_{\max}$ is the total number of iterations, iter is the current number of iterations, and n is the nonlinear adjustment factor.

3.3. Automatic Acquisition of IMF.

For single-failure faults, the sensitive IMF is automatically obtained using the principle of the maximum weighted AFM. For compound faults, the sensitive IMF is automatically obtained using the maximum SPR principle.

The SPR can effectively extract the periodic impulse components of the signal and is immune to interference from abnormal spikes in time-domain signals [30]. A larger value of SPR indicates a larger periodic impulse component of the fault signal:

$$\text{SPRI} = \frac{k \sum_{i=1}^{H} p_I(h)}{\sum_{i=1}^{K} s(k)}, \tag{18}$$

$$\text{SPRO} = \frac{k \sum_{i=1}^{H} p_O(h)}{\sum_{k=1}^{K} s(k)}, \tag{19}$$

$$\text{SPRB} = \frac{k \sum_{i=1}^{H} p_B(h)}{\sum_{k=1}^{K} s(k)}, \tag{20}$$

$$\text{SPRC} = \frac{k \sum_{i=1}^{H} p_C(h)}{\sum_{k=1}^{K} s(k)}. \tag{21}$$

Here, $s(k)$ denotes the amplitude of each point on the envelope spectrum, $k = 1, 2, \ldots, K$ denotes the number of envelope lines; $p_I(h)$, $p_O(h)$, $p_B(h)$, $p_C(h)$ denote the amplitude peaks corresponding to the fault characteristic frequencies of the estimated bearing inner ring's frequencies—that is, the bearing outer ring, roller, and cage, respectively. Parameters $h = 1, 2, \ldots, H$ denote the number of harmonic waves (for the purpose of this paper, $h = 3$).

3.4. Mutual Information.

Mutual information (MI) [31] can quantitatively represent the correlation between two random variables, and it is more accurate than the correlation coefficient method. Mutual information can effectively measure the coupling degree between IMF components. The stronger the coupling of variables is, the greater the mutual information is. It is defined as follows:

$$\text{MI}(X, Y) = \iint\limits_{s} f_{x,y}(x, y) \log \frac{f_{x,y}(x, y)}{f_x(x) f_y(y)} \, dx \, dy, \tag{22}$$

where S is the domain of X and Y, $f_{x,y}(x, y)$ is the joint distribution of X and Y, and $f_x(x)$ and $f_y(y)$ are the marginal distributions of X and Y, respectively.

3.5. Proposed Method.

A flowchart of the method proposed in this paper is shown in Figure 1. The procedure is as follows:

(1) Input vibration signal f; set the range of the VMD parameter $\gamma = (K, \alpha)$ to be optimized and initialize the IPSO parameters.

(2) Randomly initialize the particles' positions, calculate the objective function of each particle, and save the position corresponding to the maximum WAFM. That is, the parameter combination $\gamma = (K, \alpha)$.

(3) Construct a weighted AFM indicator as the objective function for optimization. Use the IPSO algorithm to update the position of the particle swarm and calculate the objective function of the particle swarm. Determine whether overdecomposition occurs using the minimum value of mutual information from each decomposed component and the original signal. If the minimum value of mutual information is less than 0.1, the current optimal solution is not updated; otherwise, the optimal solution is updated.

(4) Judge whether the stop condition is met: if so, exit to with optimal solution; if not, repeat step 3.

(5) Perform VMD on the original signal using the determined optimal parameter set $\gamma = (K, \alpha)$. For a single-failure fault, the sensitive IMF is automatically obtained using the principle of the maximum weighted AFM. For compound faults, the sensitive IMF is automatically obtained using the maximum SPR principle.
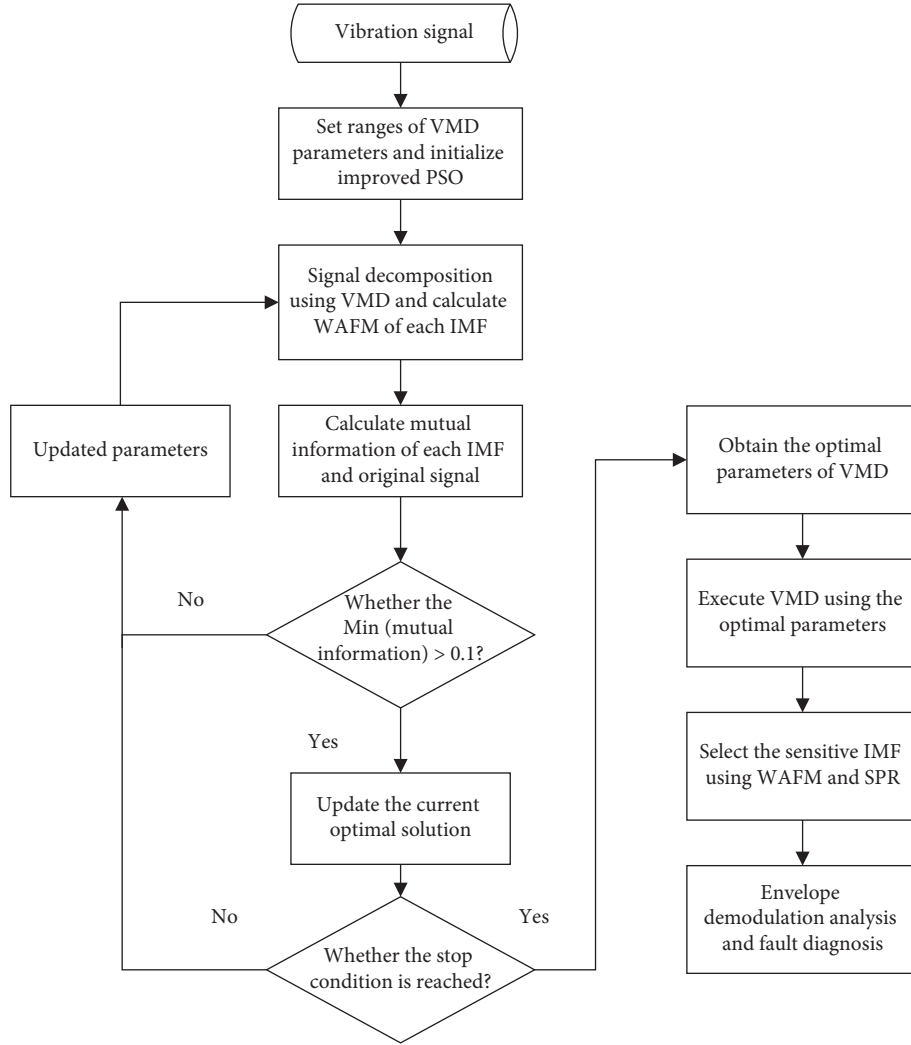
Figure 1: Flowchart of the parametric adaptive VMD method.

(6) Envelope demodulation analysis is performed on the IMF to obtain the fault characteristics.

4. Simulation Signal

In order to verify the effectiveness of the method proposed in this paper, the following model is used to simulate the vibration signal of the failure of the bearing outer ring, and add random noise with an amplitude of 2. The simulation signal is as follows:

$$y_1(t) = y_0 e^{-\xi \omega_n t} \sin \omega_n \sqrt{1 - \xi^2} t, \tag{23}$$

$$y(t) = y_1(t) + n(t), \tag{24}$$

where $y_1(t)$ denotes the impact signal of a bearing fault, an intrinsic frequency $f_n = 3000$ Hz, $\omega_n = 2\pi f_n$; a displacement constant $y_0 = 1$, a damping coefficient $\xi = 0.1$, an impact fault frequency of 100 Hz, and a sampling frequency $f_s = 20$ kHz, with a sampled point count of $N = 4096$. The waveform of the simulation signal is shown in Figure 2.

It can be seen from Figure 2 that, under the interference of random noise, the periodic pulse in the simulation signal

has been buried, and it is difficult to find the periodic impulse signal in the time-domain waveform. The envelope demodulation analysis of the simulation signal is conducted, and the resulting envelope spectrum is shown in Figure 3, in which failure characteristic components cannot be observed.

4.1. Optimization of the VMD Parameters. The simulation signal is being decomposed by the parametric adaptive VMD method proposed in this paper (search range of K: [2, 12]; search range of α: [100, 5000]). The setting of PSO parameters in this paper is obtained by adjusting according to experience and previous studies on PSO. In this paper, two parameters need to be optimized, so the dimension of the particle is 2. According to the literature [32], PSO is not sensitive to population size. At present, the study of learning factors c_1 and c_2 is limited to the application of the problem. Considering the computational complexity and reliability, the population size was set as 10, c_1 as 1.7, and c_2 as 1.5 according to the experience after several tests and adjustments. The authors of literature [33] found that when the inertia weight belongs to [0.4, 0.95], the performance of
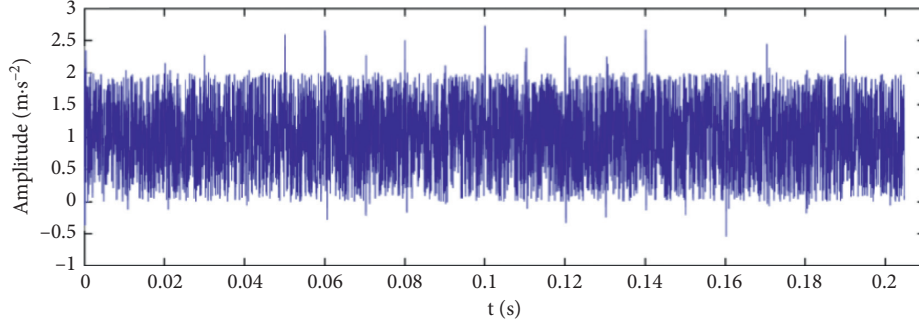
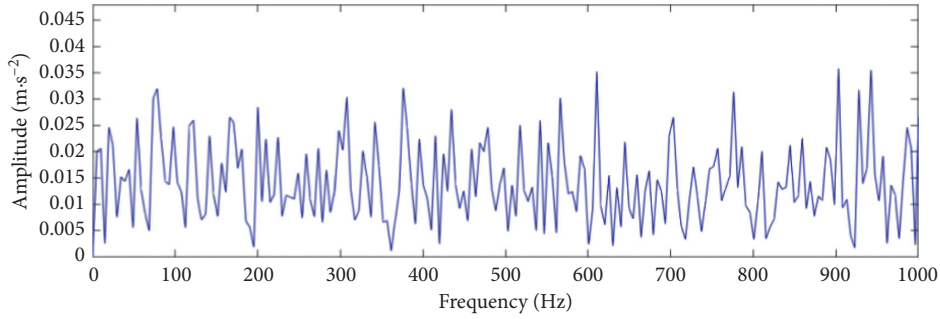FIGURE 2: Time-domain waveform of simulation signal.

FIGURE 3: Hilbert envelope spectrum of simulation signal.

particle swarm optimization algorithm will be greatly improved. So, this paper sets maximum inertia weight $w_{\max}$ = 0.95 and minimum inertia weight $w_{\min}$ = 0.4. Initialize the IPSO, and set the population size $N = 10$, particle dimension $D = 2$, maximum evolutionary generation number $t = 20$, acceleration coefficients $c_1 = 1.7$, $c_2 = 1.5$, maximum inertia weight $w_{\max} = 0.95$, minimum inertia weight $w_{\min} = 0.4$, and nonlinear modulation index $n = 1.2$ [29]. Figure 4 shows the iteration curve for the optimization search, from which approximate a global maximum max (WAFM) = 0.0782 and the optimal parameter combination of [10, 3146] can be obtained.

4.2. VMD and Sensitive IMF Selection. On this basis, the optimal parameter combination [10, 3146] is used to perform the VMD of the original signal. The waveform of each IMF component after decomposition is shown in Figure 5(a). The histogram of IMF-WAFM is shown in Figure 5(b), in which it can be seen that the WAFM value of IMF4 is the highest, so it is selected for further envelope spectrum analysis, and the effect is shown in Figure 5(c).

Envelope spectrum analysis was also performed on the other IMF components. Due to limited space, only envelope spectrums of IMF3 and IMF5 were given, as shown in Figures 5(d) and 5(e). In the envelope spectrum of IMF5 in Figure 5(e), failure characteristic components cannot be observed. Although prominent spectral line can be observed at the fault characteristic frequency f_o in Figure 5(d), there is a lot of background noise. In contrast, in the envelope spectrum of sensitive IMF in Figure 5(c), the spectral lines at

FIGURE 4: IPSO convergence curve for VMD parameter optimization.

the fault characteristic frequency f_o and double frequency $2f_o$ are more prominent. In Figure 5(c), the rotation frequency f_r can also be observed significantly. This suggests that the improved VMD method proposed in this paper can effectively decompose the signal component containing fault characteristic components from the original signal buried by noise, which proves the effectiveness of the method.

4.3. Comparison with Other Methods. In order to verify the advantages of WAFM index proposed in this paper, the same optimal combination [10, 3146] is used for VMD, and sensitive IMF is selected through the maximum kurtosis principle. It is found that IMF6 is a sensitive IMF and that no
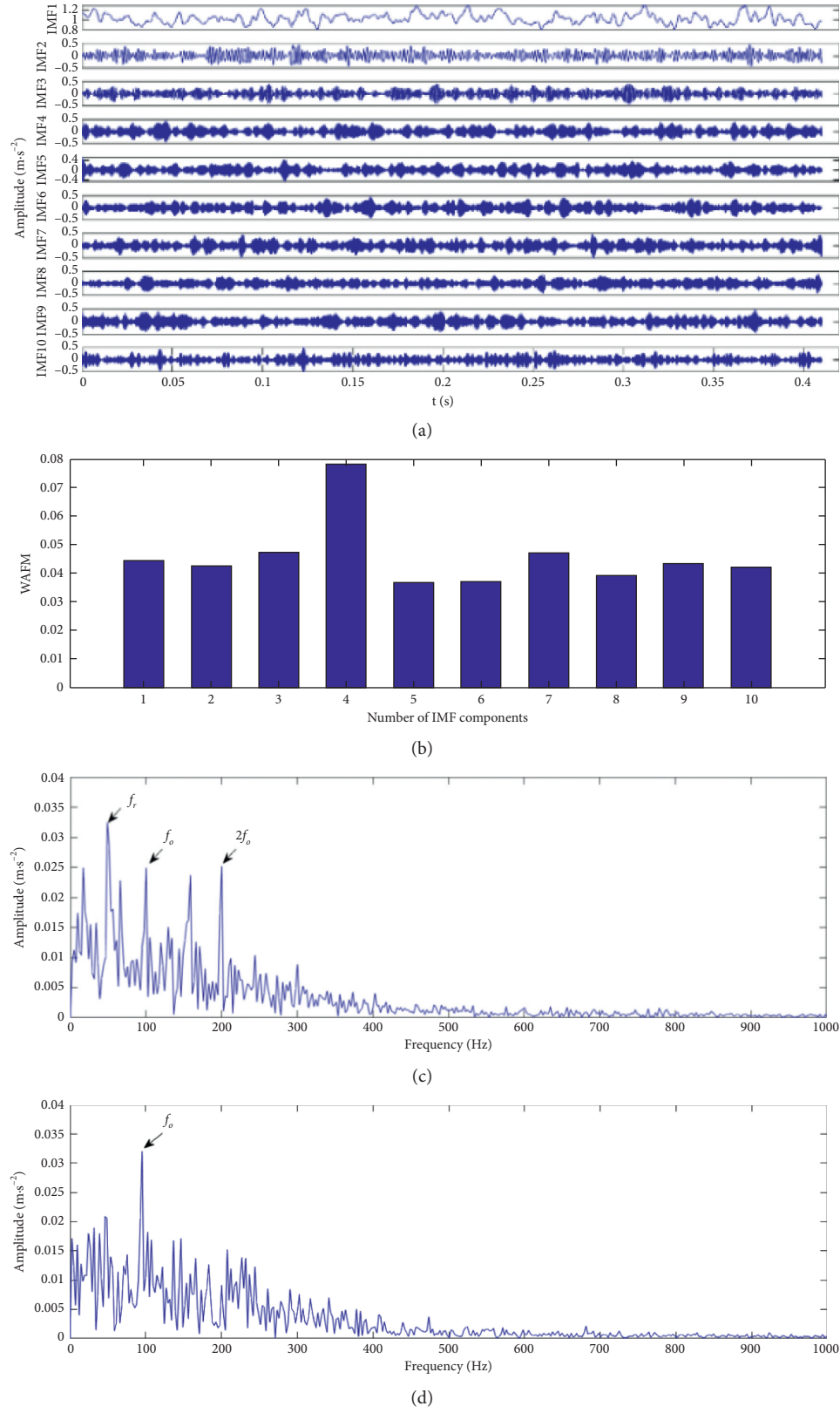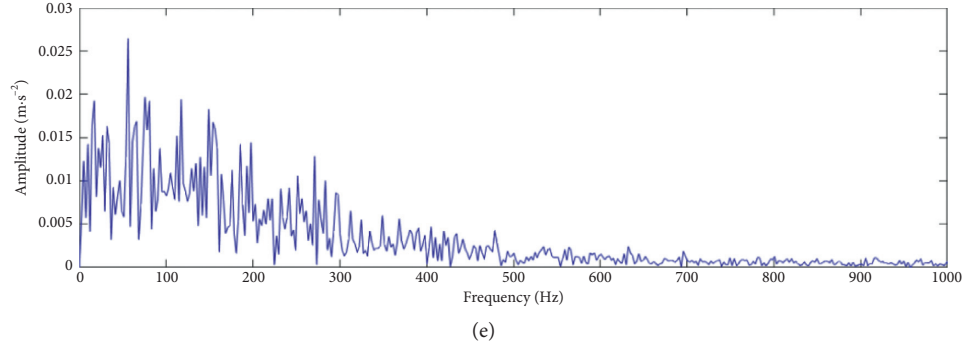
(a)

(b)

(c)

(d)

Figure 5: Continued.

(e)

Figure 5: Analysis results of simulation signal. (a) The modes obtained by the improved VMD method. (b) WAFM indexes of IMF components. (c) Envelope spectrum of IMF4 component. (d) Envelope spectrum of IMF3 component. (e) Envelope spectrum of IMF5 component.
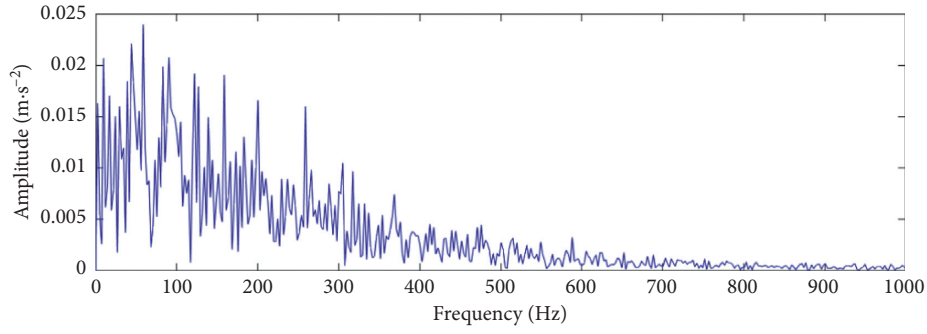

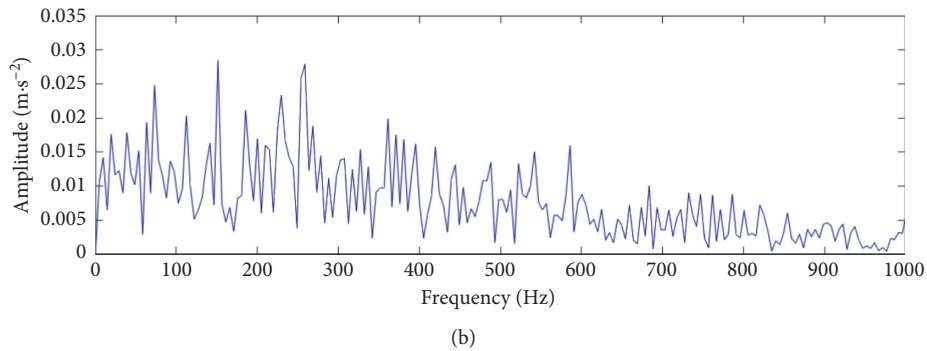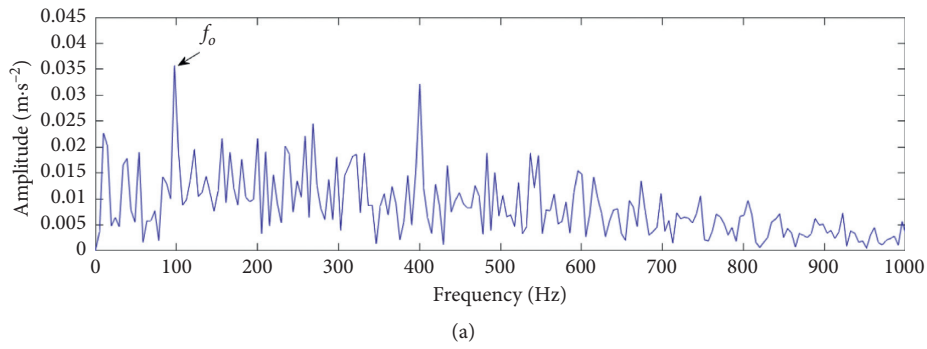Figure 6: Envelope spectrum of sensitive IMF component.

(a)

(b)

Figure 7: Analysis results of randomly selected parameter combinations. (a) Combination of parameters [10, 400]. (b) Combination of parameters [5, 3000].
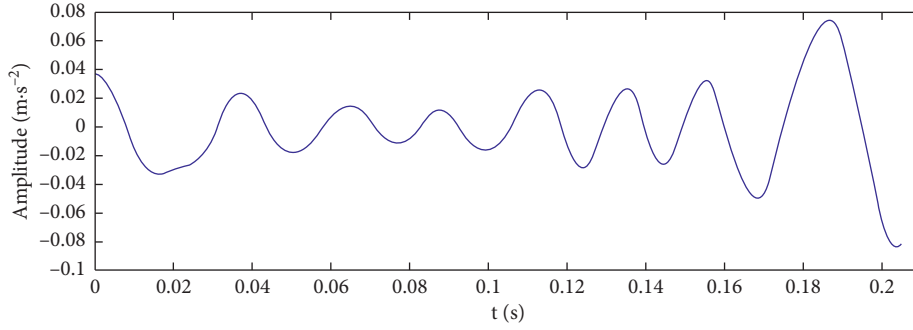
Figure 8: Time-domain waveform of sensitive IMF component by the EMD method.
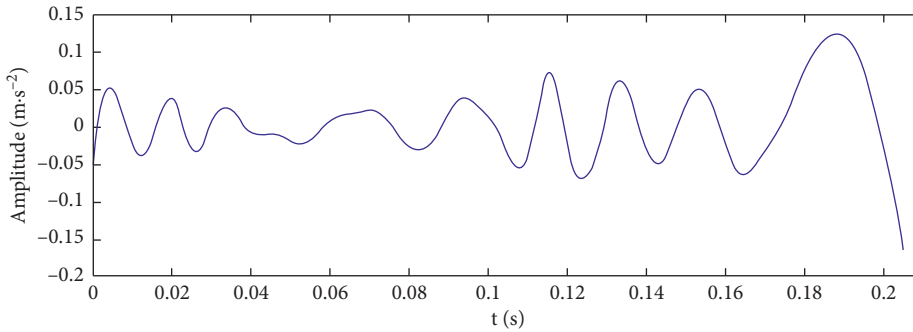
Figure 9: Time-domain waveform of sensitive IMF component by the EEMD method.

fault characteristic components can be observed from its envelope spectrum, as shown in Figure 6.

To verify the advantages of the optimization method introduced in this paper, the combinations of parameters [10, 400] and [5, 3000] are artificially selected for VMD. The sensitive IMF envelope spectrum selected with WAFM can be seen in Figure 7. According to our findings, fault characteristic components cannot be observed in Figure 7(b). The fault characteristic frequency f_o can be observed on the envelope spectrum of sensitive IMF in Figure 7(b), but it has been buried by noise. This reveals the fact that there is a gap between the method of artificial parameter selection and the method proposed in this paper.

For the purpose of further validating the advantage of the method proposed in this paper, EMD and EEMD algorithms were adopted to decompose the simulation signal. The two parameters of EEMD, ensemble number and amplitude coefficient, were set as 100 and 0.01 in [34]. After decomposition, maximum kurtosis principle is adopted to select the acquired IMFs. The time-domain waveforms of the sensitive IMFs are shown in Figures 8 and 9 [35, 36], in which fault characteristics cannot be observed. It is clear that the VMD method proposed in this paper is superior to both EMD and EMMD.

5. Experimental Signal Analysis

In order to verify the effectiveness of this method, it is applied to the actual data for analysis. The test rig is shown in Figure 10. The bearing type is HRB6304, the sampling frequency is 10 kHz, and 4096 data points are sampled every
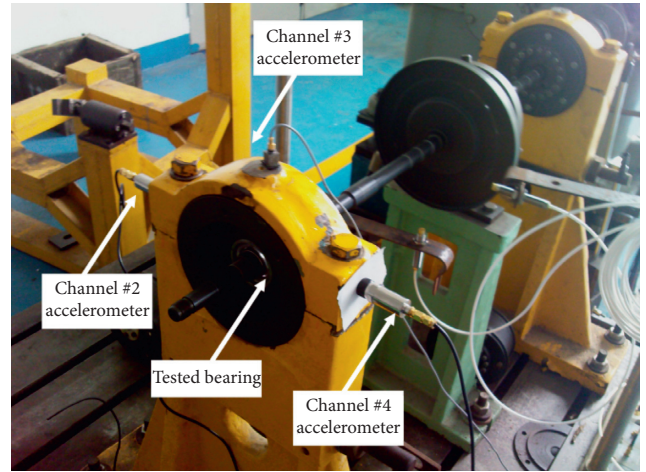

Figure 10: Test rig.

1 minute. Failure of bearing inner ring of bearing and compound failure of inner and outer ring were selected for analysis, respectively, and the failure diagrams are shown in Figure 11. Table 1 shows the multiple of fault frequency of each bearing component.

5.1. Single Fault Data Analysis. This section analyses the data of inner race fault, and the 111th minute signal of the experiment is extracted for decomposition. The velocity of the signal is at 3,000 rpm, and the bearing inner race fault characteristic frequency $f_i = 136.74$ Hz can be calculated according to Table 1. The time-domain waveform and its
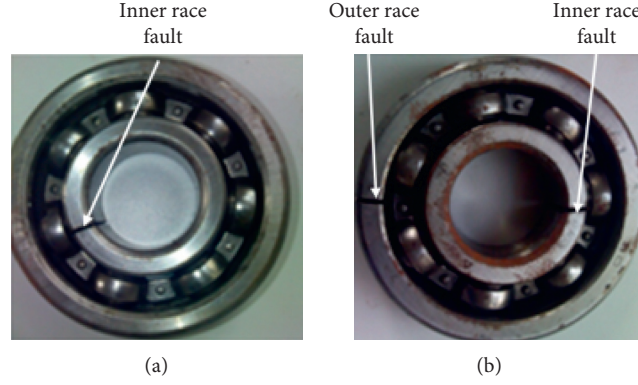
(a)

(b)

FIGURE 11: The tested bearing with fault. (a) The tested bearing with inner race fault. (b) The tested bearing with compound fault.

TABLE 1: Multiple of fault frequency of each bearing component.

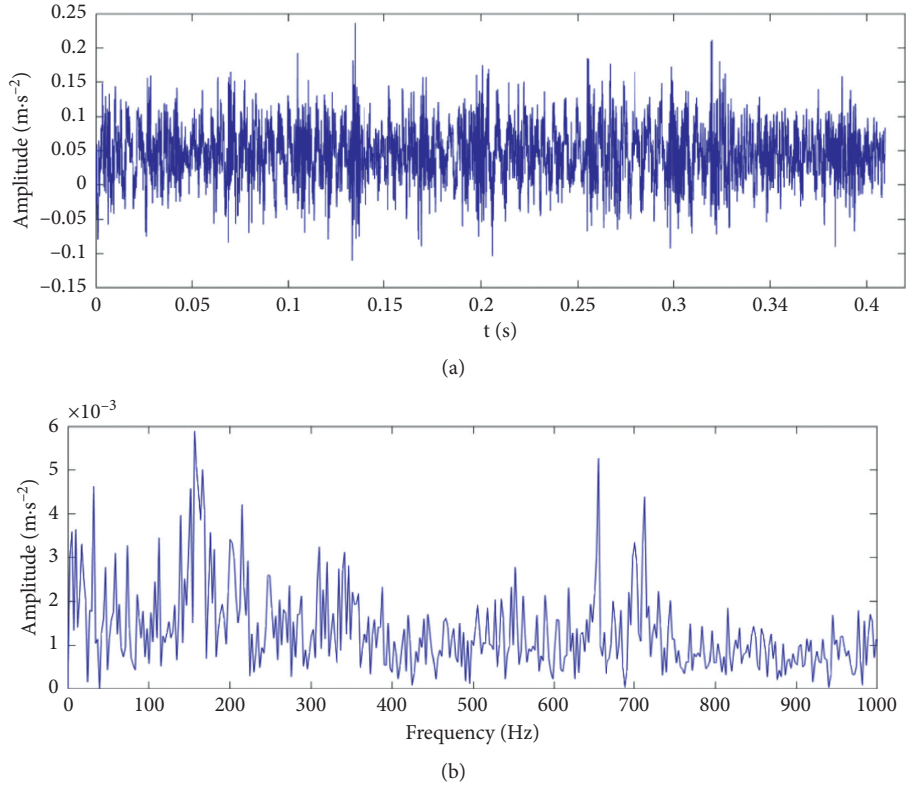| Type	Inner race	Outer race	Cage	Ball
6304	4.43	2.566	0.367	1.742

(a)

(b)

FIGURE 12: Waveform and envelope spectrum of inner race fault signal. (a) Waveform of fault signal. (b) Envelope spectrum of fault signal.

envelope spectrum of this signal are shown in Figure 12. It can be observed that no periodic impact is visible on the time-domain waveform, and there is no prominent spectral line near the fault characteristic frequency in Figure 12(b), so it is impossible to judge the state of bearing accurately.

By using the method proposed in this paper, the signal is decomposed, and the approximate global maximum max (WAFM) = 0.1356 and the optimal parameter combination

$[K, \alpha] = [10, 2416]$ can be obtained. On this basis, the optimal parameter combination [10, 2416] is used to perform the VMD of the original signal. The waveform of each IMF component after decomposition is shown in Figure 13(a). The histogram of IMF-WAFM is shown in Figure 13(b), in which it can be seen that the WAFM value of IMF8 is the highest, so it is selected for further envelope spectrum analysis, and the effect is shown in Figure 13(c).
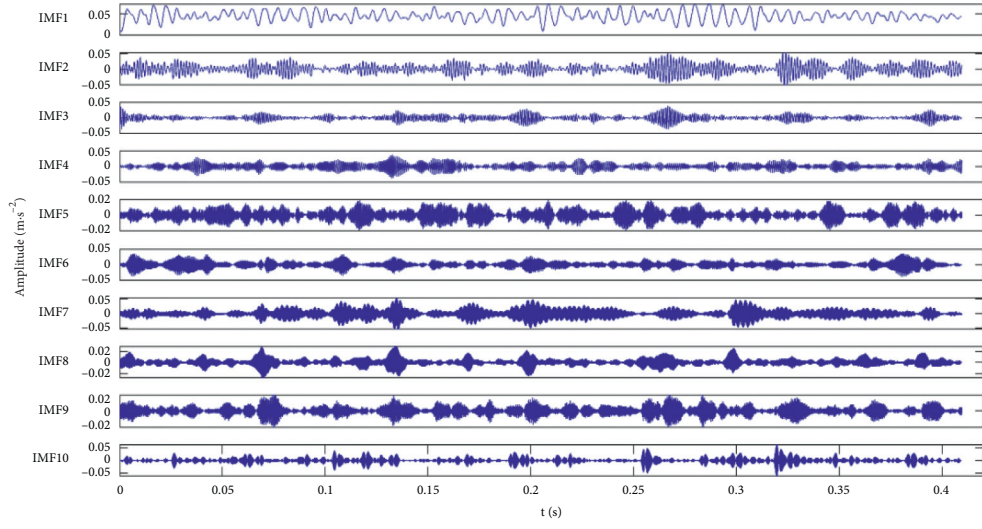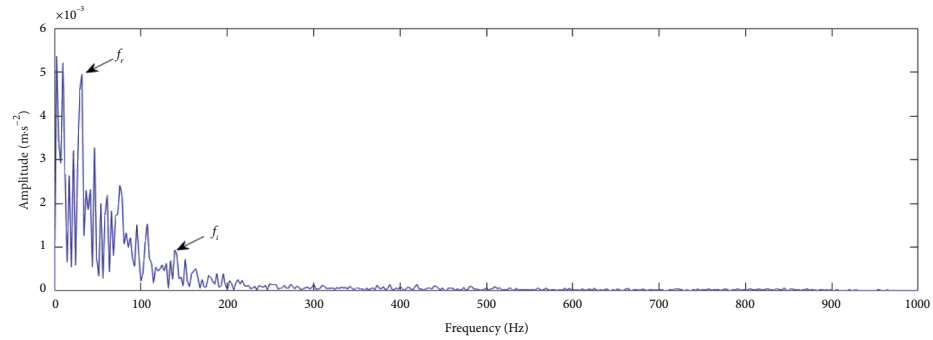
(a)
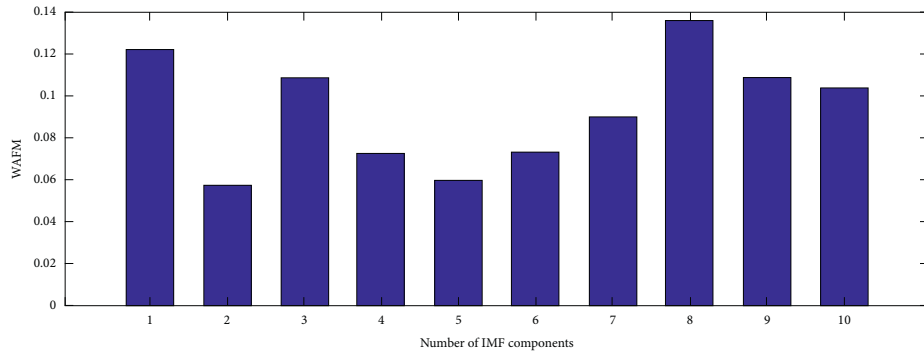
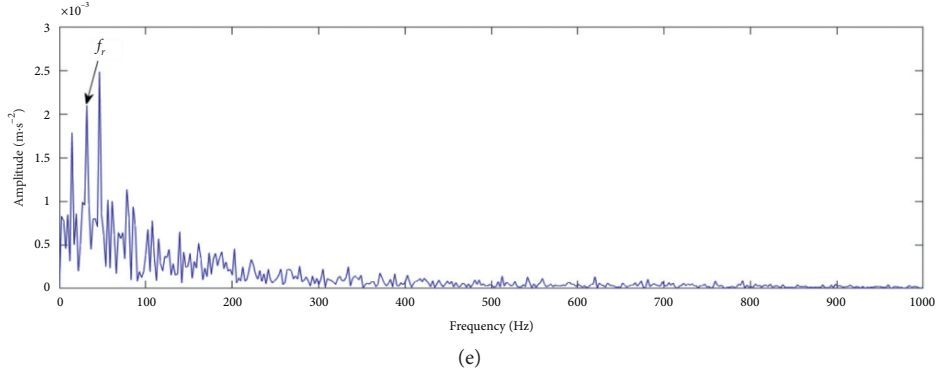

(b)
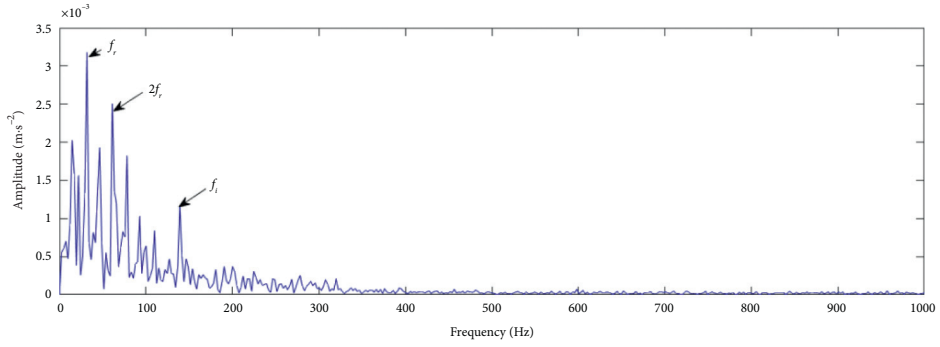
(c)

(d)

FIGURE 13: Continued.

(e)

FIGURE 13: Analysis results of inner race fault signal. (a) The modes obtained by the improved VMD method. (b) WAFM indexes of IMF components. (c) Envelope spectrum of IMF8 component. (d) Envelope spectrum of IMF7 component. (e) Envelope spectrum of IMF9 component.
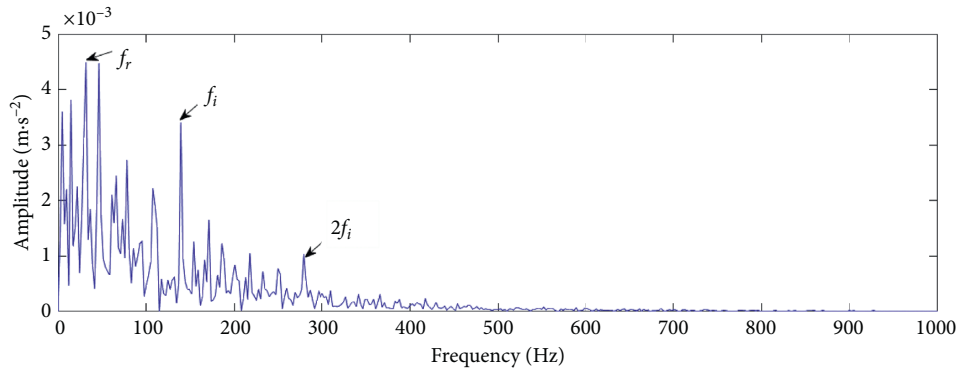

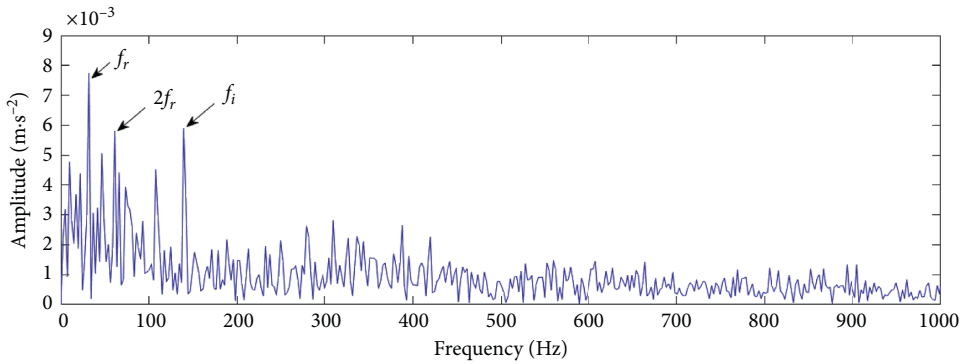FIGURE 14: Envelope spectrum of sensitive IMF component.

FIGURE 15: Envelope spectrum of sensitive IMF component by EMD method.

Envelope spectrum analysis was also performed on the other IMF components. Due to limited space, only envelope spectrums of IMF7 and IMF9 were given, as shown in Figures 13(d) and 13(e). In the envelope spectrum of IMF9 in Figure 13(e), failure characteristic components cannot be observed. The envelope spectrum of IMF9 component in Figure 13(d) shows that prominent spectral lines cannot be observed obviously at the inner race fault characteristic frequency f_i. In contrast, in the envelope spectrum of sensitive IMF in Figure 13(c), the spectral lines at the fault characteristic frequency f_i are more prominent. On this basis, the following conclusion can be drawn: the bearing inner ring has serious local defects.

Similarly, in order to verify the advantages of WAFM index proposed in this paper, sensitive IMF is selected through the maximum kurtosis principle. It is found that
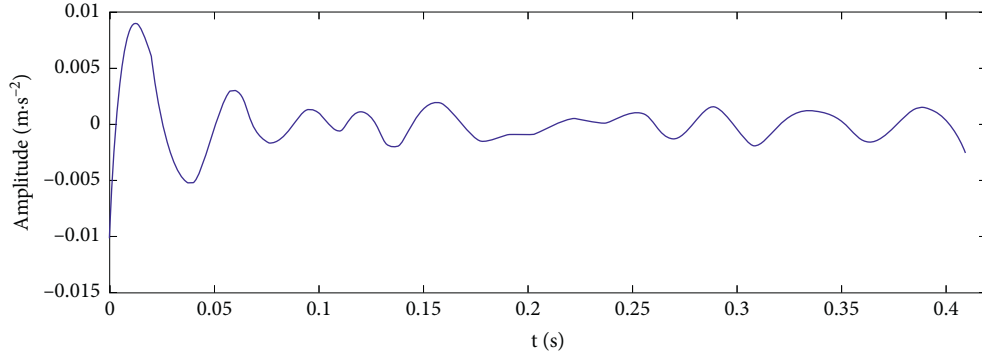
FIGURE 16: Time-domain waveform of sensitive IMF component by EEMD method.
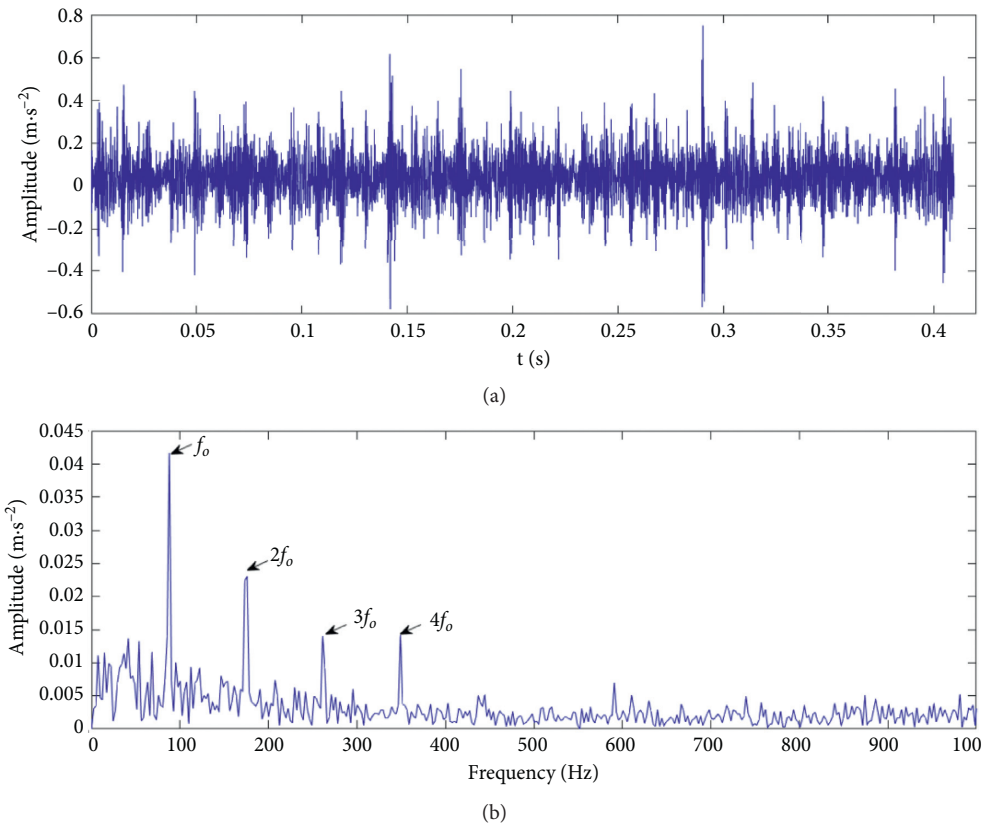

(a)

(b)

FIGURE 17: Waveform and envelope spectrum of compound fault signal. (a) Waveform of fault signal. (b) Envelope spectrum of fault signal.

IMF10 is a sensitive IMF, and its envelope spectrum is shown in Figure 14. Although the envelope line amplitudes at the fault characteristic frequency f_i and double frequency $2f_i$ can be observed, there is a lot of background noise, so it shows that the effect is not as good as that of the method proposed in this paper.

EMD and EEMD are used for signal decomposition, and kurtosis was used to select sensitive IMF. As shown by their time-domain waveform and envelope spectrum, although fault characteristic frequency f_i can be observed in Figure 15, there are more noise interference patterns in the envelope spectrum obtained in this way, so the effect is less

obvious than that in Figure 13(c). Failure characteristic components cannot be observed in Figure 16. It is, thus, clear that the VMD-based decomposition method proposed in this paper is superior to both EMD and EEMD.

5.2. Analysis of Compound Faults Data. A dataset of inner-outer ring coupling faults was used as a sample for analysis. The rotational speed indicated by the data was 2020 r/min. According to Table 1, the characteristic frequency of the inner ring fault was $f_i = 149.14$ Hz and the characteristic frequency of the outer ring fault was $f_o = 86.39$ Hz. The time
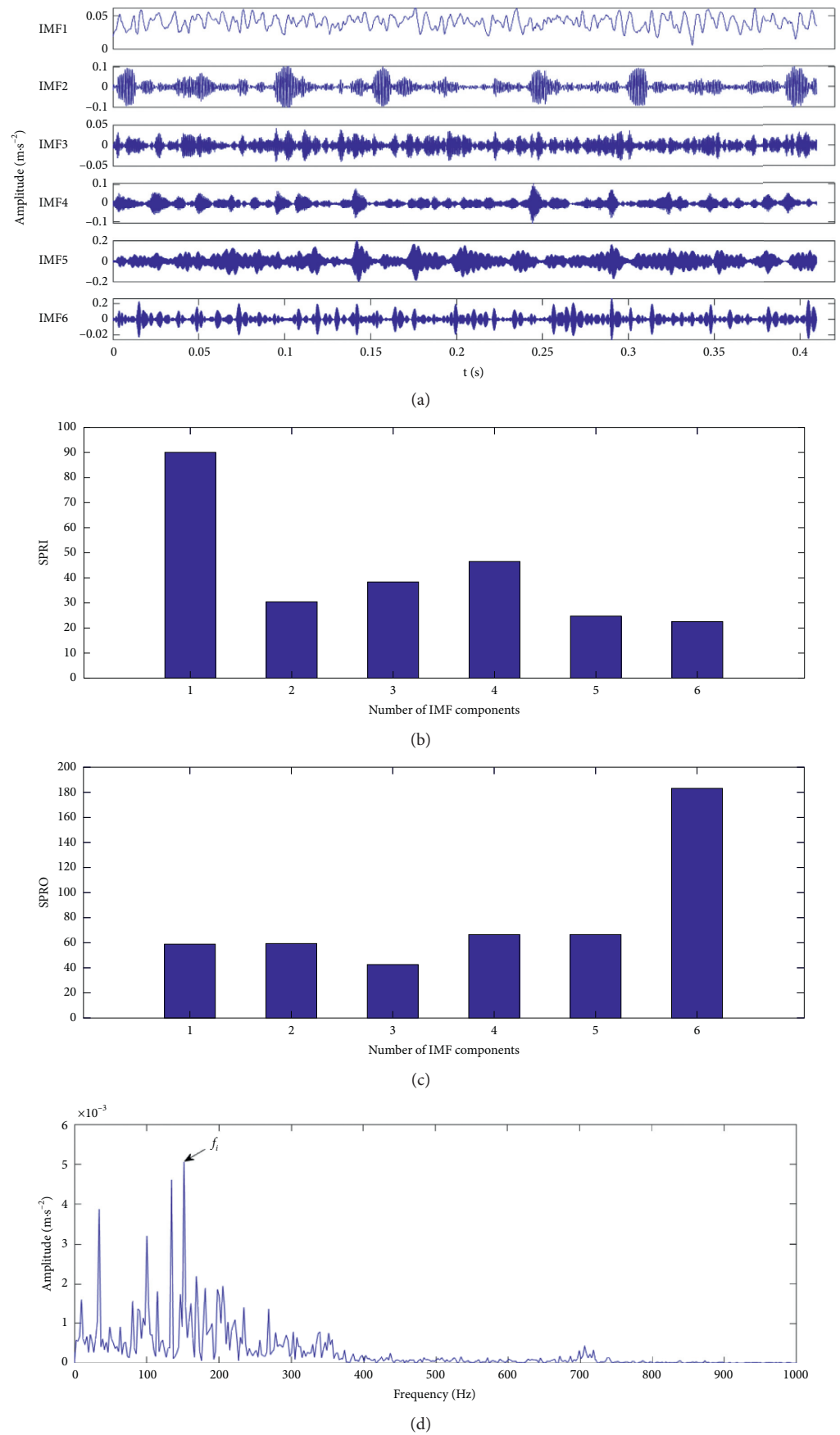
(a)
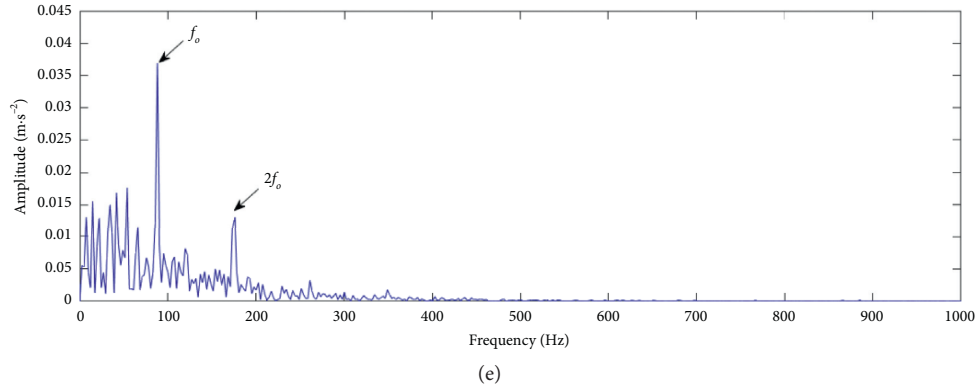
(b)

(c)

(d)

FIGURE 18: Continued.

(e)

FIGURE 18: Analysis results of compound fault signal. (a) The modes obtained by the improved VMD method. (b) SPRI indexes of IMF components. (c) SPRO indexes of IMF components. (d) Envelope spectrum of IMF1 component. (e) Envelope spectrum of IMF6 component.
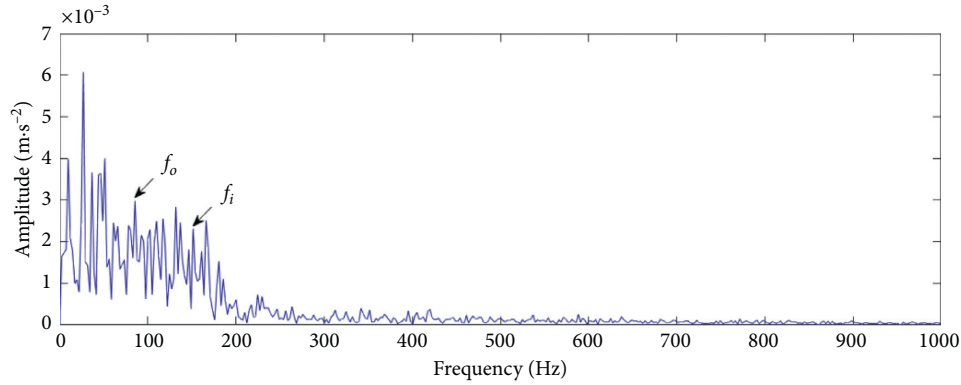

FIGURE 19: Envelope spectrum of sensitive IMF component by EMD method.

domain waveform diagram and envelope spectrum of the group of signals are shown in Figure 17. No salient periodic impact can be observed in Figure 17(a). In the envelope spectrum as shown in Figure 17(b), the characteristic of the outer ring fault can be clearly observed; however, the characteristic of the inner ring fault cannot, making it difficult to determine the bearing condition.

By using the method proposed in this paper, the signal is decomposed, and the approximate global maximum max (WAFM) = 0.2203 and the optimal parameter combination $[K, \alpha] = [6, 4605]$ can be obtained. On this basis, the optimal parameter combination is used to perform the VMD of the original signal. The waveform of each IMF component after decomposition is shown in Figure 18(a). The histograms of IMF-SPRI and IMF-SPRO are shown in Figures 18(b) and 18(c), in which it can be seen that the SPRI value of IMF1 is the highest and the SPRO value of IMF6 is the highest, so the IMFs are selected for further envelope spectrum analysis, and the effects are shown in Figures 18(d) and 18(e).

As shown in Figure 18(d), the fault characteristic frequency f_i of the inner ring can be observed in the envelope spectrum of the IMF1 component. Salient lines can be

observed near the characteristic frequency f_o of the outer ring fault and its double frequency $2f_o$ in the envelope spectrum of the IMF6 component in Figure 18(e). This indicates serious defects on both the inner and outer rings.

Signal decomposition was performed using EMD and EEMD, and the sensitive IMF was selected using SPR. After EMD processing, the sensitive IMF component obtained through SPRI-based selection or via SPRO-based selection were the same, as shown in Figure 19. The fault characteristic frequency is clearly buried in noise to the extent that no prominent spectral lines can be observed at the location of the fault characteristic frequency.

After EEMD processing, the sensitive IMF of the inner ring fault obtained via SPRI-based selection is IMF3 and the sensitive IMF of the outer ring fault obtained through SPRO-based selection is IMF5. The envelope spectrum is shown in Figure 20. No fault characteristic frequency can be observed in the sensitive IMFs of the inner ring and outer ring, indicating that the VMD-based decomposition method proposed in this paper is superior to EMD and EMMD.

In order to verify the advantage of SPR index, this paper uses WAFM maximum principle, AFM maximum principle, and kurtosis maximum principle to select sensitive IMFs.
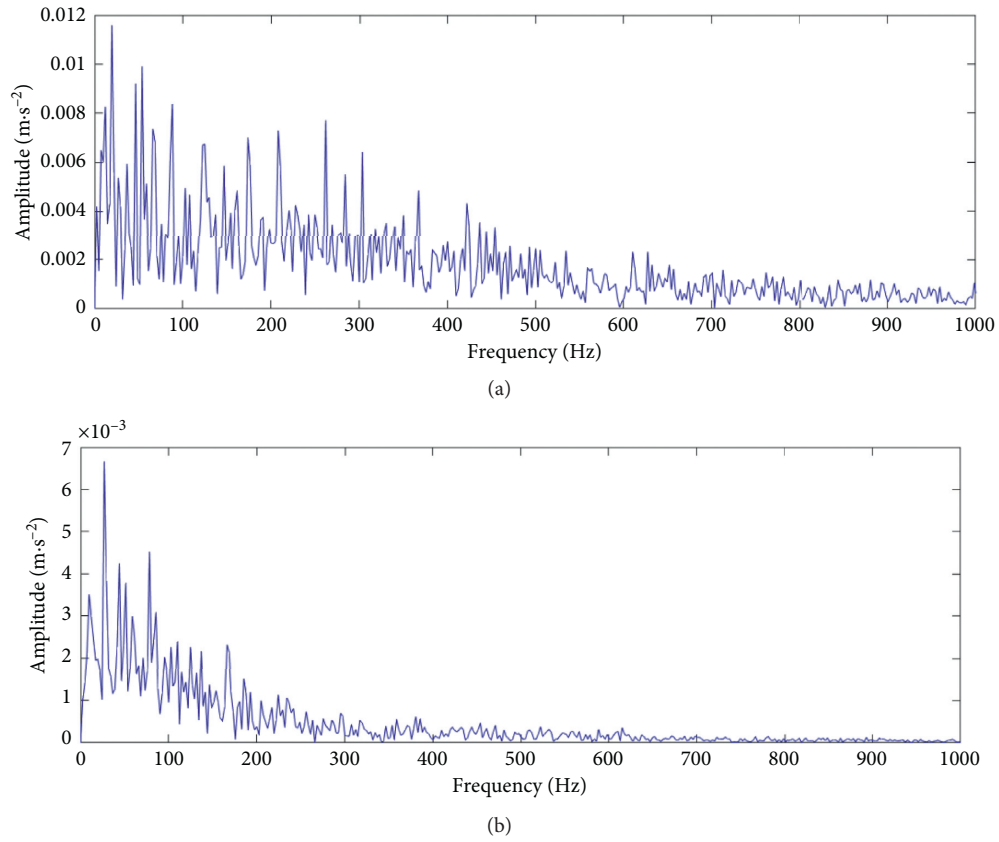
(a)

(b)

Figure 20: Envelope spectrum of sensitive IMF component by EEMD method. (a) Sensitive IMF component of inner race fault. (b) Sensitive IMF component of outer race fault.
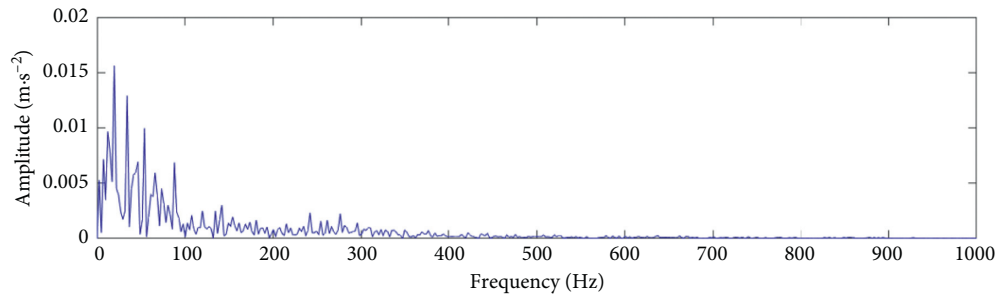

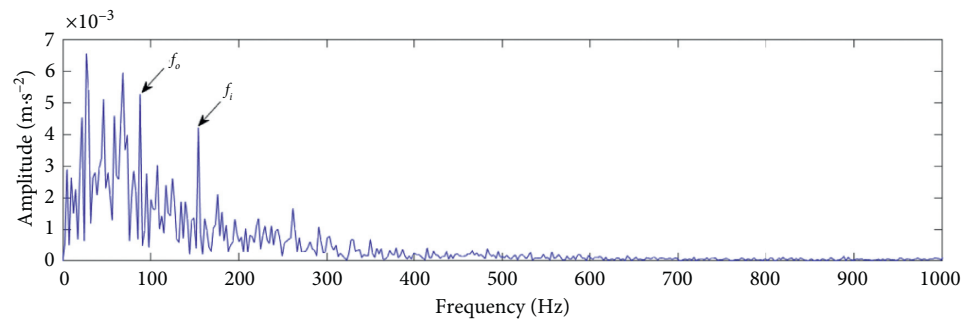Figure 21: Envelope spectrum of IMF2.

Figure 22: Envelope spectrum of IMF4.

The two sensitive IMFs with the maximum WAFM value selected by WAFM maximum principle and AFM maximum principle are IMF2 (Figure 21) and IMF6 (Figure 18(e)), respectively. IMF6 can observe the fault characteristic frequency of the outer ring, which is the same as the IMF selected by SPR, while IMF2 cannot observe the fault characteristic frequency.

The two sensitive IMFs with kurtosis maximum principle are IMF2 (Figure 21) and IMF4 (Figure 22), respectively. IMF2 cannot observe the fault characteristic frequency, and the fault frequency of inner and outer ring can be found in the envelope spectrum of IMF4, but there are many interferences, and the effect is not as good as that of the method in this paper.

6. Conclusions

(1) This paper proposes a method for improving VMD and automatically acquiring the sensitive IMF. The proposed strategy solves the problem of setting parameter combination for the conventional VMD method.

(2) The weighted AFM indicator constructed in this paper takes into account the impulsive and periodic nature of fault signals and makes it possible to search for the optimal VMD parameter set using the IPSO algorithm. Moreover, the proposed method has the benefit of adaptive parameter acquisition and the decomposition result was found to be superior compared to that of other methods such as EMD and EEMD.

(3) In view of the differences between single-failure faults and compound faults, this paper proposes using two different principles (principle of maximum weighted AFM and principle of maximum SPR) for selecting the sensitive IMF, thereby solving the difficult problem of accurately detecting compound faults. The proposed method could potentially provide technical support for engineering applications.

Data Availability

The data for this article are available from the public dataset website: http://ides.nuaa.edu.cn.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] Y. Lei, L. Jing, Z. He et al., "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems & Signal Processing*, vol. 35, no. 1-2, pp. 108–126, 2013.

[2] N. E. Huang, Z Shen, S. R. Wu et al., "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[3] Y. Lei, Z. He, and Y. Zi, "Application of the EEMD method to rotor fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1327–1338, 2009.

[4] J. Cheng, Y. Y. Yang, and Y. Yang, "A rotating machinery fault diagnosis method based on local mean decomposition," *Digital Signal Processing*, vol. 22, no. 2, pp. 356–366, 2012.

[5] J. Zheng, J. Cheng, and Y. Yang, "Partly ensemble empirical mode decomposition: an improved noise-assisted method for eliminating mode mixing," *Signal Processing*, vol. 96, pp. 362–374, 2014.

[6] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.

[7] P. Shi and W. Yang, "Precise feature extraction from wind turbine condition monitoring signals by using optimized variational mode decomposition," *IET Renewable Power Generation*, vol. 11, no. 3, 2016.

[8] X.-B. Wang, Z.-X. Yang, and X.-A. Yan, "Novel particle swarm optimization-based variational mode decomposition method for the fault diagnosis of complex rotating machinery," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 1, pp. 68–79, 2018.

[9] X. Yan, M. Jia, and L. Xiang, "Compound fault diagnosis of rotating machinery based on OVMD and a 1.5-dimension envelope spectrum," *Measurement Science and Technology*, vol. 27, no. 7, Article ID 075002, 2016.

[10] X. Zhang, H. Q. Miao, and L. Wang, "A parameter-adaptive VMD method based on grasshopper optimization algorithm to analyze vibration signals from rotating machinery," *Mechanical Systems and Signal Processing*, vol. 108, pp. 58–72, 2018.

[11] C. Liu, C. L. Zhu, and C. Ni, "Chatter detection in milling process based on VMD and energy entropy," *Mechanical Systems and Signal Processing*, vol. 105, pp. 169–182, 2018.

[12] J. Lian, H. Z. Liu, and X. Dong, "Adaptive variational mode decomposition method for signal processing based on mode characteristic," *Mechanical Systems and Signal Processing*, vol. 107, pp. 53–77, 2018.

[13] H. Zhao and L. Li, "Fault diagnosis of wind turbine bearing based on variational mode decomposition and teager energy operator," *IET Renewable Power Generation*, vol. 11, no. 4, 2016.

[14] J. Xingxing, L. Shunming, and C. Chun, "A novel method for adaptive multiresonance bands detection based on VMD and using MTEO to enhance rolling element bearing fault diagnosis," *Shock and Vibration*, vol. 2016, Article ID 8361289, 20 pages, 2016.

[15] L. Shangkun, T. Guiji, W. Xiaolong et al., "Time-frequency analysis based on improved variational mode decomposition and teager energy operator for rotor system fault diagnosis," *Mathematical Problems in Engineering*, vol. 2016, Article ID 1713046, 9 pages, 2016.

[16] H. Te and J. Dongxiang, "Rolling bearing fault diagnostic method based on VMD-AR model and random forest classifier," *Shock and Vibration*, vol. 2016, Article ID 5132046, 11 pages, 2016.

[17] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics- a tutorial," *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 485–520, 2011.

[18] O. Janssens, R. Schulz, V. Slavkovikj et al., "Thermal image based fault diagnosis for rotating machinery," *Infrared Physics & Technology*, vol. 73, pp. 78–87, 2015.

[19] Y. Miao, M. Zhao, and J. Lin, "Improvement of kurtosis-guided-grams via gini index for bearing fault feature identification," *Measurement Science and Technology*, vol. 28, no. 12, 2017.

[20] D. Chen, Y. J. Lin, and Y. Li, "Modified complementary ensemble empirical mode decomposition and intrinsic mode functions evaluation index for high-speed train gearbox fault diagnosis," *Journal of Sound and Vibration*, vol. 424, pp. 192–207, 2018.

[21] Z. Wang, J. Wang, and W. Du, "Research on fault diagnosis of gearbox with improved variational mode decomposition," *Sensors*, vol. 18, no. 10, p. 3510, 2018.

[22] Z.-X. Yang, X. Wang, and P. K. Wong, "Single and simultaneous fault diagnosis with application to a multistage gearbox: a versatile dual-elm network approach," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 12, pp. 5245–5255, 2018.

[23] J. H. Zhong, J. J. Y. Liang, Z. X. Yang et al., "An effective fault feature extraction method for gas turbine generator system diagnosis," *Shock and Vibration*, vol. 2016, Article ID 9359426, 9 pages, 2016.

[24] Z. Wang, J. Wang, W. Cai et al., "Application of an improved ensemble local mean decomposition method for gearbox composite fault diagnosis," *Complexity*, vol. 2019, Article ID 1564243, 17 pages, 2019.

[25] Z. Wang, W. Du, J. Wang et al., "Research and application of improved adaptive momeda fault diagnosis method," *Measurement*, vol. 140, pp. 63–75, 2019.

[26] Z. Wang, J. Zhou, J. Wang et al., "A novel fault diagnosis method of gearbox based on maximum kurtosis spectral entropy deconvolution," *IEEE Access*, vol. 7, pp. 29520–29532, 2019.

[27] H. Dalton, "The measurement of the inequality of incomes," *The Economic Journal*, vol. 30, no. 119, pp. 348–361, 1920.

[28] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95—International Conference on Neural Networks*, IEEE, Perth, WA, Australia, August 2002.

[29] A. Chatterjee and P. Siarry, "Nonlinear inertia weight variation for dynamic adaptation in particle swarm optimization," *Computers & Operations Research*, vol. 33, no. 3, pp. 859–871, 2006.

[30] Y. Lei, Y. Z. He, and X. Chen, "New clustering algorithm-based fault diagnosis using compensation distance evaluation technique," *Mechanical Systems and Signal Processing*, vol. 22, no. 2, pp. 419–435, 2008.

[31] L. Paninski, "Estimation of entropy and mutual information," *Neural Computation*, vol. 15, no. 6, pp. 1191–1253, 2003.

[32] Y. Shi, "A modified particle swarm optimizer," in *Proceedings of the IEEE ICEC Conference*, Singapore, December 1998.

[33] Y. H. Shi and R. C. Eberhart, "Empirical study of particle swarm optimization," in *Proceedings of the Congress on Evolutionary Computation-CEC*, IEEE, Honolulu, HI, USA, August 2002.

[34] Y. Zhang, F. H. Zuo, and F. Bai, "Classification of fault location and performance degradation of a roller bearing," *Measurement*, vol. 46, no. 3, pp. 1178–1189, 2013.

[35] A. Parey, F. El Badaoui, and N. Tandon, "Dynamic modelling of spur gear pair and application of empirical mode decomposition-based statistical analysis for early detection of localized tooth defect," *Journal of Sound and Vibration*, vol. 294, no. 3, pp. 547–561, 2006.

[36] Y. Lei, Y. Z. He, and Y. Zi, "EEMD method and WNN for fault diagnosis of locomotive roller bearings," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7334–7341, 2011.